Food Waste and the Sharing Economy

Timothy J. Richards and Stephen F. Hamilton*
Paper presented at ASSA Annual Meeting
Philadelphia, PA January 5, 2018

December 30, 2017

Abstract

Wasting food is one of the rare problems that affects our ability to achieve economic goals in terms of food security, environmental sustainability, and farm-financial security. Most of the ideas proposed to this point involve either behavioral nudges or administrative regulations that are either too paternalistic or piecemeal to represent viable solutions. In this study, we investigate the potential for commerical peer-to-peer mutualization systems (CPMSs), or sharing-economy firms, to emerge as market platforms for the exchange of surplus food. If a system of CPMSs is able to develop in a self-sustaining way, then the market prices they create will generate sufficient incentives for all actors to manage surplus food more efficiently. We develop an empirical model of a CPMS operating as a platform in a two-sided market, and examine its viability using data from one of the first CPMS firms in the surplus-harvest industry, Imperfect Produce, Inc. Empirical estimates of a two-sided network-demand model show that user-demand rises in the number of growers shipping to the platform, and grower demand for distribution rises in the number of users. Our findings indicate that secondary markets have the key elements needed for CPMS success, and that policy tools designed to facilitate transactions in secondary markets can be highly effective in reducing food waste.

*Richards is Professor and Morrison Chair of Agribusiness, Morrison School of Agribusiness, W. P. Carey School of Business, Arizona State University; Hamilton is Professor and Department Head, Economics Department, Orfalea College of Business, California Polytechnic State University, San Luis Obispo, San Luis Obispo, CA. Contact author: Richards, Address: 7231 E Sonoran Arroyo Mall, Mesa, AZ, 85212. Ph. 480-727-1488, email: trichards@asu.edu. Funding from Agriculture and Food Research Initiative (National Instittue for Food and Agriculture, USDA) grant no. 2016-09900 is gratefully acknowledged.

1 Introduction

Food waste is one of the rare problems that cuts across multiple social issues, from food security (Coleman-Jenson, et al. 2014) and environmental degradation to economic efficiency (Parfitt, et al. 2010; Gustavsson, et al. 2011; Buzby et al. 2011; Buzby and Hyman 2012). Wasted food not only impairs society's ability to feed an estimated 9.7 billion people globally by 2050 (UN 2015), but it also accounts for roughly 25% of US freshwater supplies each year and consumes nearly 300 million barrels of oil (Hall et al. 2009). Food production generates substantial environmental externalities associated with greenhouse gas emissions and phosphate run-off (Buzby and Hyman 2012), the unconsumed portion of which is unnecessary, and food waste at the terminal point of the food system acounts for roughly 18% of total solid waste in municipal landfills (EPA 2016). In terms of the discarded value of food, alone, USDA estimates that the US loses 31% of total food supply, or $165.5 billion per year in total value (Buzby, et al. 2014). Food waste occurs at virtually all stages of the supply chain from farmer to retailer to consumer, resulting in the disposal of potentially usable food in nearly every sector of the food system in the distribution channel between farmers and consumers.

An important strand of economic research examines consumer food waste as a behavioral problem, seeking to address the problem by regulating waste disposal by educating consumers about expiration dates and changing consumers' incentives to generate waste (Tsiros and Heilman 2005; Theotokis, et al. 2012; Buzby and Hyman 2012; Halloran et al. 2014).[1] These are important priorities that confront a growing population. However, policies designed to reduce food waste in the consumer market only control incentives at the end of the

[1]Emerging research has identifed a range of causes specific to different levels of the food-supply chain. While losses at the farm level due to weather damage and natural variation in quality are substantial (Gustavsson et al. 2011; Kummu et al. 2012), most of the waste in developed economies comes from households (Griffin et al. 2009; Buzby et al. 2011; Cicatiello et al. 2016). Food waste at the household level is primarily due to a lack of understanding of "best-before" or "use-by" dates, inaccurate meal planning, imperfect home-storage systems, and discounts on large packages that encourage over-buying (Gustavsson et al. 2011; Halloran et al. 2014). Demand uncertainty, and the inability to accurately forecast demand, are also key to food waste among foodservice operators and food manufacturers (Mena et al. 2011), resulting in the so-called bullwhip effect that magnifies food waste through the supply chain (Lee at al. 1997; Sucky 2009).

distribution channel and fail to encompass all stages of the food system where food waste occurs. In the upstream stages of the food economy, commercial peer-to-peer mutualization systems (CPMSs), that seek to match farmers and distributors to consumers for fresh produce items, represent a potentially important market-based solution to more efficiently allocate food at higher levels of the food sytem by stimulating price realization for products that are edible, but contain defects in size, color, shape and size; the so-called market for "ugly food" (Garfield 2016).[2] In this paper, we examine the efficiency of food exchange in upstream markets of the food system by examining the performance of CPMS systems in facilitating the exchange of harvested produce that is too small, large, misshapen, or discolored to make saleable grade through traditional marketing channels.

It has long been understood in the management of negative environmental externalities that attaching a price to an activity is more likely to lead to its efficient control, including waste (Dinan 1993; Fullerton and Kinnaman 1995; Fullerton and Wu 1998; Buzby and Hyman 2012; Acuff and Kaffine 2013). By creating a market for imperfect, or surplus, food, suppliers will be better able to match the distribution of quality produced by the natural variability of biological production, with the willingness-to-pay for quality in the consumer market.[3] This is a classic price discrimination result – whereas supermarket grading standards (which are generally higher than USDA grades for fresh produce) serve as an effective minimum-quality standard, selling imperfect produce that is inarguably below-grade allows suppliers to segment consumers according to their willingness to pay for quality (Mussa and Rosen 1978; Caswell 1998), sell a greater quantity, and reduce the amount of surplus-harvest. In this paper, we investigate the potential for CPMS markets to emerge for such ugly food products.

Fresh foods that are harvested below marketable quality and left unsold contribute to

[2] We view CPMS systems broadly as any technology that matches buyers and sellers on a platform derived from peer-to-peer transactions. Botsman (2013) provides a general characterization of CPMS firms as any entity that facilitates the decentralized trade of products or services that are underutilized in the economy.

[3] A reviewer suggests that creating markets for surplus harvest will increase supply, reduce the price, and encourage more waste. While this effect is plausible, we believe it represents a second-order effect relative to the direct incentive embedded in a price for waste.

the food waste problem. Indeed, Gustavsson et al. (2011) and Kummu et al. (2012) have found weather damage and natural variation in quality to lead to substantial quantities of unsalable farm products. Because farmers face uncertain demand and supply conditions, and forecasting is imperfect, over-production of farm products occassionally occurs, leading to low price realizations that leave perfectly edible food unharvested in the field. In the absence of a market for surplus, or leftover food, excessive amounts of usable food are discarded either by being "plowed under" at the farm level or by being harvested and sent to downstream markets that may ultimately process these products as waste (Garrone, et al. 2014). CPMS services that help match these products with buyers can offer an important market for reducing food waste.[4]

We base our observations on on the performance of CPMS systems for surplus food on data from Imperfect Produce, Inc. Imperfect Produce, Inc. is a startup company based in California that aims to reduce food waste in the surplus-harvest market by matching producers at the farm level with consumers at the retail and foodservice levels of the food system for the exchange of food products that are not graded through conventional channels of the food system. Our data consists of four years of peer-to-peer transactions, including the amounts ordered, prices paid, and attributes of consumers and the sharing firms. These data are sufficiently rich to allow us to test an empirical model of activity on the sharing platform in which the breadth of sales transactions matches with the range of consumer preferences for product attributes that drive value in final goods markets for food products. These data are sufficiently rich to allow us to investigate whether the fundamental conditions are present for a CPMS to succeed in matching buyers and sellers in food market, and if so, whether farm-to-consumer platforms in the "food sharing" economy present a viable opportunity for a upstream food markets to help alleviate the problem of unwanted food.[5]

[4]In this study, we do not attempt to measure the amount of food loss, or waste, so we use the terms interchangably throughout. We do appreciate, however, that the concepts are not identical (Bellemare, et al. 2017).

[5]While food waste occurs at all points of the food supply chain, CPMSs to date have emerged largely between farmers and consumers, providing rich transactional data on which to base our empirical investigation. However, Food Cowboy represents one example of a "B2B" firm that transacts surplus food that has

CPMS firms such as Uber, AirBnB, FarmLink, TaskRabbit, and Liquid have increased consumer's willingness to transact goods in the "sharing economy" (Bardhi and Eckhardt 2012; Lamberton and Rose 2012; Sundararajan 2013, 2014; Belk 2014; Fraiberger and Sundararajan 2015; Möhlmann 2015). As Botsman and Rogers (2010) argue, CPMS markets emerge when advances in sharing technology – e.g., cell phone applications – facilitate markets for durable assets with excess capacity. In the case of food, a farmer's field is the durable asset, and excess capacity is manifest in surplus harvest. A novel feature of CPMS markets for surplus food is that excess capacity in food markets results in a perishable stock. For this reason, policies that facilitate food transactions in a CPMS market, and thereby generate sales that would otherwise not transpire, serve to reduce surplus output that would otherwise be discarded, plowed under, or end up in lower-valued uses than intended.[6]

Our empirical approach is framed around recent estimation techniques employed in two-sided markets (Armstrong 2006; Kaiser and Wright 2006; Steiner, et al. 2016). In a two-sided market, demand for a "platform", for instance a menu of food items coordinated for sale by a CPMS provider like Imperfect Produce, is comprised by demand for distribution from potential suppliers of surplus food on one side, and by demand for procurement from potential consumers of food waste on the other side. The nature of demand on the platform is two-sided due to indirect network economies (Rochet and Tirole 2003, 2006) created by the breadth of the items available on the platform. Specifically, the benefit to consumers from interacting on the platform rises with the number of suppliers providing surplus food

been purchased by restaurants, or even households. In this regard, Food Cowboy represents an example of how the CPMS concept may be extended to downstream food markets and encompass consumer-level food waste.

[6] An interesting possibility suggested by a reviewer is that creating new, upstream markets for surplus harvest will reduce food prices, potentially resulting in greater post-consumer food waste. While our model is silent on general equilibrium effects, the total amount of food transacted on CPMS platforms is currently small and likely to have negligible effects on overall food prices. Moreover, we believe enhancing the efficiency of food utilization at upstream levels of the food system leads to better matching in downstream consumer markets that will tend to dominate second-order effects relatived to changes in consumer food prices. Comprehensive modeling of general equilibrium effects of CPMS innovations in upstream food markets, which includes changes in land use, changes in animal feed prices, as well as consumer price changes in fresh produce markets relative to processed, shelf-stable foods (and the attendant consumer health implications), are beyond the scope of the present study.

on the platform, while the benefit to suppliers from interacting on the platform rises with the number of consumers purchasing surplus food on the platform. Network economies on a two-sided platform thereby create a "virtuous cycle" in which supply facilitates its own demand, causing emerging platforms either to succeed or fail in spectacular fashion.

We estimate the strength of demand on each side of the market in our empirical model, allowing us to determine: (i) whether demand conditions exist for CPMS markets to emerge as a viable business model for surplus food; and (ii) whether policy tools such as subsidies on "ugly food" are effective in reducing the amount of surplus food. Our findings indicate that consumers' preferences for the breadth of food items available on the site is particularly important in driving indirect network effects in the CPMS market. This result suggests that the profitability of a CPMS in this setting is directly related to the number of suppliers the platform sources from in procuring surplus food, and that the value of transacting surplus food on the platform rises significantly for producers with the size of the network. This feature of the market makes food policies that subsidize purchases on CPMS platforms for surplus food particularly effective in reducing excess produce. To quantify this effect, we numerically estimate the effect of price subsidies on CPMS purchases on the quantity of surplus food siphoned out of the waste stream in upstream food markets. We find that a 25 percent subsidy on CPMS transactions results in a 60 percent increase in the quantity of surplus food sold on the platform and that a 90 percent subsidy on CPMS transactions results in a threefold increase in the quantity of surplus food sold on the platform, which suggests that subsidies on "ugly food" transactions are a cost-effective policy to reduce the amount of food that may be otherwise discarded.

Our study contributes both to the empirical literatures on two-sided markets and CPMS viability and to the practical policy discussion on food waste. While the majority of empirical investigations into two-sided markets consider either technology (Nair, Chintagunta, and Dube 2004) or media (Ackerberg 2004; Kaiser and Wright 2006) markets, ours is the first to consider a secondary market for a surplus commodity as a fundamentally two-sided market.

That is, consumers demand a range of choices from the commodity on offer, and suppliers demand broad distribution among the consumers served by the platform. Our empirical analysis of a surplus-harvest CPMS is the first to cast the study of sharing-economy firms in the context of a two-sided market, with the attendant implication that they may be subject to indirect network effects. In terms of the food waste literature, ours is the first analysis of the viability of a "market" for surplus food. If such markets are indeed viable, then agents at all points in the food supply chain may be endowed with the incentive to not throw food away, but to trade it for profit.

The remainder of the paper is organized as follows. In the next section, we formulate an empirical model to assess the viability of CPMS platforms as a tool for reducing surplus food. In section 3, we describe our data and provide summary evidence from reduced-form models that the market for surplus food indeed has the character of a two-sided market. In section 4, we present and discuss our estimation results that test our hypothesis of a two-sided market and discuss the practical relevance of our findings for the viability of CPMS platforms for creating markets for surplus harvest. In the final section, we draw some broad conclusions on how policies can be designed to reduce surplus food through the use of targeted incentives in emerging CPMS markets for ugly produce.

2 Econometric Model of a Surplus-Food Market

2.1 Background

Our empirical model is based on the indirect network effects generated by a two-sided demand for intermediary distribution services.[7] CPMSs in general, and those that distribute surplus food specifically, behave as multi-product platforms that exist to connect buyers of surplus food to suppliers. Buyers prefer a variety of products, and therefore value platforms that are able to attract and retain a large number of suppliers (Draganska and Jain 2005; Richards

[7] "Indirect network effects" refer to the general concept that benefits of membership rise in the "size" of the market, whether measured by the number of users, products, software titles or other measure.

and Hamilton 2013; Steiner, et al. 2016), while suppliers prefer to sell through a platform that attracts a large number of potential buyers. In the case of a CPMS market for surplus food, buyers may be more interested in procuring food from the CPMS market when the platform makes multiple food items available at once, and sellers can find higher-valued uses for surplus food items when there is a greater number of buyers on the platform.

Network effects have been empirically identified in many markets for durable goods. Network effects have been shown to be important in two-sided markets for computer hardware and software (Nair, Chintagunta, and Dube 2004), video games (Clements and Ohashi 2005; Corts and Lederman 2009; Dube, Hitsch, and Chintagunta 2010; Lee 2013; Zhou 2016), automated clearing house (ACH) payment systems (Ackerberg and Gowrisankaran (2006), intermediation systems (Caillaud and Jullien 2003), video cassette recorders (Park 2004), compact-disc players (Gandal, Kende, and Rob 2000), C2C platforms (Chu and Manchanda 2016), radio stations (Jeziorski 2014), sports-card trading platforms (Jin and Rysman 2015), newspapers (Argentesi 2007; Chandra and Collard-Wexler 2009; Van Cayseele and Vanormelingen 2009), yellow page advertising and magazines (Rysman 2004, Kaiser and Wright 2006). Our analysis departs from the literature by considering network effects for perishable items (surplus produce), which differ methodologically from durable goods markets (as discussed below), and by deriving novel insights for alternative uses for unsold food. Specifically, our approach allows us to numerically simulate the effectiveness of various food policy instruments (e.g., taxes on waste, subsidies on donations) in reducing surplus harvest when food policy is targeted toward one side or the other of the CPMS market.

Our objective in this section is to derive a model of weekly platform demand, where demand is defined as the probability of ordering food on the platform, multiplied by the size of each order. The size of each order is defined as the number of different items purchased in each basket.

We adopt a two-stage approach to modeling each component of platform demand. In the first stage, we model the probability that each household purchases surplus food from the platform in a given week, and in the second stage, we model the number of items purchased on the platform conditional on having placed an order. Aggregating over all households in the data set accordingly provides a predictive model of the total platform demand on a weekly basis. Critically, platform demand in each stage of the model depends on the range of products offered, or the number of suppliers serving the platform.

With the demand estimates, we then model equilibrium product provision, or how suppliers of surplus food respond to consumer demand conditions on the platform. By endogenizing both demand for the platform, and the provision of surplus food on the platform, we estimate the strength of demand on each side of the market, which identifies the importance of indirect network effects in the CPMS market for surplus food.

2.2 Purchase Incidence Model

We begin by modeling the number of orders transacted on the platform each week, which is the product of the total number of households visiting the platform and the probability that each household purchases food items on the platform. The probability that an individual household places an order (purchase incidence) depends, in turn, on the variety of items offered on the platform and the prices of the various items. Allowing household purchase incidence to depend on the variety of products available on the platform captures the indirect network effect in which demand depends on the number of surplus food items available on the platform. That demand depends on the extent of product variety available in the marketplace is well-established in both the marketing literature (McAlister and Pessemier 1982; Kim, Allenby and Rossi 2002; Briesch, Chintagunta and Fox 2009) and the economics literature (Dixit and Stiglitz 1977; Richards and Hamilton 2015). Intuitively, if preferences are distributed uniformly among consumers in the market, then individual consumers are more likely to find better matches between products and their tastes when a greater number of products is available, raising the probability of finding an acceptable match on the platform.

We model our first stage "Order-Probability" demand using a combined constant elasticity of substitution (CES) - logit framework. Consumers form expectations of the number of items offered on the platform, and the total sales receipt that results from their subsequent purchase, according to a CES index function; however, the indirect utility of each order is assumed to be driven by an Extreme Value (logit) preference-heterogeneity assumption. The CES form ensures that consumers exhibit an inherent preference for variety (Dixit and Stiglitz, 1977; Nair, Chintagunta and Dube, 2004) when forming their expectations of surplus food purchases on the CPMS platform, which is appropriate for our order-demand problem because utility is assumed to rise in the number of items available, and at the same time does not restrict the degree of substitution among products as with a discrete-choice model.

We start with a brief description of the Order-Probability model, and then discuss how we nest the Order and Size models together on the CPMS platform. Assume the platform offers $j = 1, 2, ...N$ products, where a product is defined as an item in one of several categories, such as fresh fruits, vegetables, cereal products, or other perishable items.[8] Assume buyer i visits the platform and obtains utility from buying products $j = 1, 2, ... N$ during week t as given by the CES demand model (suppressing the time, t, subscript to simplify notation):

$$U_i(q_{i1}, q_{i2}, ...q_{iN}, z_i) = \left(\sum_{j=1}^{N} q_{ij}^{\theta} \right)^{\sigma} + z_i, \tag{1}$$

where the value within parentheses, $Q_i = \sum_{j=1}^{N} q_{ij}^{\theta}$, is defined as a CES quantity index, q_{ij} is the quantity of product j purchased by consumer i, z_i is the outside or numeraire product, and $0 < \theta < 1$, and $0 < \sigma < 1$ assure concavity of the utility function. The parameter θ ensures that the products are not perfect substitutes so buyers can, but do not have to, purchase positive amounts of each product (Nair, Chintagunta and Dube, 2004). This parameter also ensures that the model generates a positive utility of variety in equilibrium. With direct utility defined over products, the buyer chooses the quantity of each subject to the usual budget constraint, with income y_i, such that the inverse demand for the products

[8] Note that we cannot assume the products within a particular category are identical over time as suppliers only deliver what happens to be in surplus at each point in time.

offered by the platform during any particular week is written:[9]

$$p(q_{i1}, q_{i2}, ...q_{iN}, y_i) = \sigma\theta \left(\sum_{j=1}^{N} q_{ij}^{\theta} \right)^{\sigma-1} q_{ij}^{\theta-1}, \tag{2}$$

We then solve for the direct demand system and substitute the result back into the utility function in (1) to find the indirect utility for buyer i choosing the bundle of N products across all categories j as:

$$V_i(p, N, y_i) = (1-\sigma\theta)(\sigma\theta)^{\frac{\sigma\theta}{1-\sigma\theta}} N^{\frac{\sigma(1-\theta)}{1-\sigma\theta}} p^{\frac{\sigma\theta}{\sigma\theta-1}} + y_i, \tag{3}$$

where y_i is the amount of the numeraire good, assuming prices for all products within each category for the platform are symmetric ($p = p_j \forall j$), so the CES price index simplifies to: $p = N_j^{\frac{\theta-1}{\theta}} p_j$. We then use this CES price index to capture the effect of basket-level pricing and the size of the platform on the probability that each household places an order during a given week.

We assume that order-preferences are heterogeneous and are randomly distributed over consumers, which allows indirect utility to be written as:

$$V_i(p, N, y_i) = (1-\sigma\theta)(\sigma\theta)^{\frac{\sigma\theta}{1-\sigma\theta}} N^{\frac{\sigma(1-\theta)}{1-\sigma\theta}} p^{\frac{\sigma\theta}{\sigma\theta-1}} + y_i + \varepsilon_i, \tag{4}$$

where ε_i is an iid random error term. Assuming the distribution of consumer heterogeneity is Type I Extreme Value, and incorporating the fact that our data is time-series in nature, the probability that buyer i purchases from the platform at time t is given by: $P_{it} = \Pr(V_{it} > V_{it}^* + \varepsilon_{it}) = \exp(V_{it})/(1 + \exp(V_{it}))$, which results in a familiar logit model of purchase-incidence, with non-linear utility (Bucklin and Lattin 1992; Bell and Lattin 1998; Briesch, Chintagunta and Fox 2009). Based on economic principles, therefore, the probability that a household purchases on the platform in a given week is a decreasing function of the average price of the basket of items he or she purchases, and an increasing function of the number of items available on the platform.

[9]Some of Imperfect Produce's buyers are foodservice operators, so the budget constraint does not apply to all. However, management assures us that these buyers are in the minority, but in number and in volume. Therefore, a consumer-based model is an accurate description of their marginal buyer.

Our data are collected at the individual-household level, and reflect specific purchases of surplus food baskets ("boxes"). As a result, both box-attributes and household preference heterogeneity are likely to be important in determining the probability that each household purchases a particular box of food from the platform in a given week. If either of these features of our data are not taken into account, they are likely to induce substantial bias in all model parameters. Because the quantity of the numeraire good drops out in finding the logit-probability term, we write the indirect utility function V_{it} for estimation purposes as:

$$V_i(p, N, \mathbf{z}, \mathbf{x}) = f(p, N|\sigma, \theta) + \sum_{j=1}^{J} \gamma_j z_{ij} + \sum_{k=1}^{K} \beta_k x_k + \varepsilon_i, \tag{5}$$

where $f()$ is the non-linear function of prices and platform size implied by the CES consumer-preference model above, $\mathbf{z}$ is a vector of household-attributes that reflect a household's need-based motivations for purchasing from the site in a given week, while $\mathbf{x}$ captures attributes of the specific box of food items that is purchased.

We calculate a number of need-based variables in $\mathbf{z}$ that are commonly used to explain purchase incidence, or the probability that a purchase occurs during a particular week (Bell, Ho, and Tang 1998; Briesch, Chintagunta, and Fox 2009). Namely: CR_i = consumption rate, or the average apparent rate of fruit and vegetable consumption per household, calculated by dividing total purchases over the sample period by the number of weeks the household participates on the site,[10] ITT_i = inter-purchase time, or the number of weeks between the previous purchase and current purchase, LQ_i = lagged quantity, or the number of items purchased on the previous purchase occasion.

Among our household-level attributes, we expect CR to have a positive influence on the probability of purchase, all else constant, as heavier fruit and vegetable consumers are likely to be more frequent visitors to the site. We expect ITT to have a similar, positive, effect on the probability of purchase because, at a given consumption rate, the longer time between purchases implies a greater likelihood that the household will run out of its preferred

[10] For all households, the relevant sample period consists of only those weeks between the first- and last-purchase weeks. Implicitly, therefore, we assume the households is not aware of Imperfect Produce prior to their first purchase, and choose not to use the site after their final purchase.

items. Lagged quantity (LQ), on the other hand, is expected to have a negative effect on the probability of purchase as stockpiling, to the extent that it is possible with perishable items, reduces the likelihood of need during the next purchase occasion.

In addition to these measures of household-heterogeneity, we include a set of basket attributes ($\mathbf{x}$) that consists of: $PROM$ = the dollar value of any promotion used to purchase a particular box of food by household i at purchase occasion t, ORG = a binary indicator that assumes a value of 1 if the contents of the box are organic, FR = a binary indicator that assumes a value of 1 if the box consists entirely of fruit, VG = a binary indicator that equals 1 if the basket consists entirely of vegetables, SM = a binary indicator that equals 1 if the box is a "small" size, MD = a binary indicator that equals 1 for "medium" size boxes, and $LG = 1$ if the box is "large". For the categorical variables, the base case for basket-content is a "mixed" box of food items, while an extra-large basket serves as the base case for the size indicator.

In the Imperfect Produce data, we have no measures of observed heterogeneity at the household level (i.e., demographic variables such as age, income, and education) so we can only control for unobserved preference heterogeneity at the household level. We control for unobserved heterogeneity by allowing the key model parameters to vary randomly across households such that:

$$\begin{aligned} \sigma_i &= \sigma_0 + \sigma_1 \upsilon_1,\ \upsilon_1 \tilde{}\, N(0,1) \qquad (6) \\ \theta_i &= \theta_0 + \theta_1 \upsilon_2,\ \upsilon_2 \tilde{}\, N(0,1), \end{aligned}$$

where a 0 subscript indicates the parameter-mean, and 1 its standard deviation, and υ_k are independent, standard-normal, random variates. Because the logit expression no longer has a closed form with parameters that vary randomly, we estimate the entire Order-Probability / Order-Size model using simulated maximum likelihood using the algorithm described below.

2.3 Order-Size Model

In the second-stage model, "Order-Size" demand is represented by a count-data framework, which is appropriate because our data describes the number of items (size) purchased with each order. Because size-demand is a weakly positive, count-variable, we model the size of each order as a Poisson-distributed variable (Bucklin, Gupta, and Siddarth 1998).[11] A count-data specification is appropriate for this problem because it is reasonable to expect that once a buyer has decided to purchase, the number of items is determined by the household's need for each item, the relative price of each, and a number of unobserved factors specific to each household.

Order-size, or the amount purchased, is estimated conditional on the observation that the buyer visits the platform. On the Imperfect Product platform, buyers have the option of purchasing either a small, medium, large, or extra-large box of food, with the number of items in each box increasing accordingly.[12] While buyers likely have many other alternatives for their fresh produce, our data do not describe their other purchases, so we focus only on the items chosen from the platform.[13] The number of items purchased in each of $j = 1, 2, ..., J$ different box formats depends on both the realization of utility for buyer i and the set of observed box-specific attributes described above (x_k) and unobserved household-preference attributes. Unlike the purchase-incidence model above, the arguments of the purchase-frequency model are intended to capture volume-preference rather than need-based measures. Assuming the number of items purchased is Poisson distributed, therefore, implies that the probability of purchasing Q_{ijt} items from the platform in period t is given by:

$$P(Q_{ijt} = q_{ijt}|Q_{ijt} > 0) = \frac{\exp(-\lambda_i)(\lambda_i)^{q_{ijt}}}{(1 - \exp(-\lambda_i))q_{ijt}!}, \tag{7}$$

where λ_i is the Poisson distribution parameter with: $\lambda_i = \exp(\phi_{i0} + \phi_p p + \phi_N N + \sum_{k=1}^{K} \phi_k x_k)$ to ensure that the visitation probability is strictly positive. Conditional on having chosen

[11] We describe our tests for overdispersion in the Estimation and Identification section below.

[12] Box sizes vary over time, and with the items chosen, so box size and number of items are not isomorphic.

[13] With data from more than one platform, the platform choice model could easily be extended to a nested logit framework (Nair, Chintagunta, and Dube 2004).

to make a purchase, the quantity purchased is likely to decrease in prices (p) because box prices rise approximately linearly with item prices. Also conditional on purchase-incidence, we expect the number of items chosen to rise in the number of items available on the platform (N) simply because consumers are assumed to have a preference for variety, and there will be a higher probability of a preference-match the more items are available on the site.[14] The remaining elements of the box-attribute vector ($\mathbf{x}$) are the same as in the logit purchase-incidence model, and we have relatively obvious priors on how each of these variables is likely to affect the number of items purchased relative to the base case scenarios. Namely, promotion will be positively related to the number of items purchased, organic items are likely to be purchased more often, and smaller-sized boxes will imply a lower number of items. As in the logit model, we include a measure of unobserved heterogeneity in the λ_i expression that is again assumed to be normally distributed.

Although the data generating process is, conceptually, maintained to follow a Poisson distribution, empirical applications of the Poisson model often find that the data are more disperse than the maintained distribution would suggest. Practical causes of overdispersion include settings in which contagion are potentially important, or bandwagon effects, in which a rise in the number of observations of a phenomenon is likely to also be associated with a greater dispersion about the mean. Intuitively, user-networks are likely to succeed, or fail, in spectacular fashion. Consequently, we begin by estimating a base Poisson model, and then consider Negative Binomial (NB) alternative as a means of addressing any overdispersion problem that may arise. We test for the preferred specification using the Chi-square test developed by Cameron and Trivedi (1990). In this test, the null hypothesis is that the mean of the estimated distribution is equal to the variance, while in the alternative, the variance is greater than the mean (hence the term overdispersion). The CT test for overdispersion

[14] Variety is discovered through the shopping experience, and not advertised explicitly. Shoppers are allowed to customize their box by first choosing the size of the box, then choosing whether it contains only fruit, only vegetables, or a mixture of the two, and can then choose individual elements of the box. More variety implies a greater ability to substitute items in and out of the box, very similar to a traditional shopping experience.

essentially involves conducting an ordinary least squares regression of the variance of the fitted value of the dependent variable on either the mean, or the square of the mean (Greene 2010). The resulting test statistic, for the significance of the regression coefficient, is Chi-square distributed with 1 degree of freedom.

In our application, we fail to reject the null hypothesis that the variance of the estimated value of box-size is equal to the mean, but reject the null if we define the alternative as the square of the mean. Consequently, we have some support for the maintained hypothesis that box-size follows a Poisson distribution, but this support is not entirely conclusive. Therefore, we present results from both the NB model below, and the simpler Poisson alternative.[15]

As in the Order-Probability model above, we also allow for unobserved heterogeneity over buyers by allowing critical elements of the Poisson λ_{ijt} function, namely prices and network-size, to be randomly distributed over buyers. Failing to account for unobserved heterogeneity in a household-level environment such as ours invites bias in all parameters. Following the notational convention introduced above, we allow each parameter to be normally distributed such that the parameters are given by: $\phi_{ik} = \phi_{0k} + \phi_{1k}\upsilon_{3k}$, $\upsilon_{3k}\tilde{}N(0,1)$ where the υ_{3k} are independent, standard-normal, random variates.

2.4 Estimation and Identification

We estimate both stages of the platform-demand model together using maximum likelihood. Combining the Order-Probability and Order-Size models, the parameters of the fixed-parameter version of the Logit-Poisson platform-demand model are estimated by maximizing the log-likelihood function value given by:

$$LLF = \sum_{T} \ln[P_{ijt}(j) * P(Q_{ijt} = q_{ijt}|Q_{ijt} > 0))^{d_{ijt}}(1 - P_{ijt}(j))^{(1-d_{ijt})}], \tag{8}$$

where $d_{ijt} = 1$ if buyer i purchases box j from the platform on visit t, and $= 0$ otherwise. For the random-parameter version of the platform-demand model, we use simulated

[15] We present the likelihood function for the Negative-Binomial "P" model in the appendix below, which is a general version of the more usual Negative Binomial model.

maximum-likelihood (SML) to recover both the structural parameters, and the distributional parameters of the random elements of the model. As is standard in this literature, we use Halton draws to improve the efficiency of the estimation routine, and found that there was little difference in the parameter estimates for more than 100 Halton draws. The result of this model is demand for both the platform and the surplus food purchased on the platform.

Before estimating equation (8) and the supply of surplus-harvest equation, it is first necessary to account for the fact that the number of products and the price index are likely to be endogenous. Therefore, we estimate the platform-demand model using a control function approach (Petrin and Train 2010), using raw commodity and other operating input prices as instruments for retail prices, and commodity shipment levels as instruments in the demand equation. Specifically, the USDA reports a full set of wholesale prices for fresh fruits and vegetables in California on a weekly basis, so these farm-gate prices constitute clear candidates for retail-price instruments. Wholesale prices for a range of fresh fruits and vegetables are the primary cost-element for Imperfect Produce, so are likely to be closely related to the retail prices they charge users. However, because Imperfect Produce is an infinitesimally small player in the US fresh produce industry, the prices they pay to farmers are plausibly unrelated to the demand for produce in general. Therefore, wholesale produce prices are likely to be excellent retail-price instruments. We also include a non-linear time trend to account for any movements in retail prices that may be driven by the same, more general, cycles that drive Imperfect Produce's pricing strategy.

A first-stage regression of retail prices set by Imperfect Produce on indices of fruit and vegetable prices, and time-trends, produces an F-statistic of $10,564.36$ and a R^2 value of 0.18. Based on these estimates, we can safely conclude that our wholesale price indices are not weak instruments for retail prices (Staiger and Stock 1997), and economic reasoning suggests that are likely to be appropriate as well.

For network size, we face a fundamental dimensionality problem in identifying the effect of network size on platform demand. Namely, over the data period IP offers some 450 different items for sale through either their vegetable, fruit, or mixed boxes. And, our prices are reported on a per-box basis and not a per-item basis. Therefore, recording and matching item-level instruments is neither feasible, nor desirable. We capture wholesale movements in the produce offered for sale by IP by using indices of fresh fruit and vegetable prices reported by USDA for the state of California (USDA 2017). Our instrument for network size exploits the uncertainty of agricultural production, and the economic rationale for establishing food-based CPMSs. That is, IP exists in order to create a market for surplus produce. In few other industries is there a greater difference between planned and actual production output, because yields and grades are largely determined by environmental conditions such as heat, rain, or wind, and not necessarily by conscious management decisions. These features of the biological production process mean that observed production levels for specialty crops are exogenous to the demand for the IP platform, and yet correlated with the number of items that appear on the site.

Our instrument consists of an index of commodity movement volumes using the USDA Agricultural Marketing Service Market News Service website. For the 60-week sample period, we created a straight sum of weekly shipments in the Southern California district (in 10,000 lb units) for the top 20 commodities that appear on the IP website. Because of the wide range of specific items offered on the site (an average of some 450 products), tracking every one of them through the AMS site is intractable. However, we assume that higher volumes of the most popular items are correlated with production, and shipment levels, of all speciality items. That is, overproduction for the most common commodities will be correlated with overproduction of all commodities within a single growing season. Our index, therefore, is expected to be positively correlated with the amount of surplus growers find on their hands, and the range of items that appear for sale on the site. First-stage IV regressions using this index find an R^2 of 0.66 and an F value of $125,553.09$, so our network-size instruments again cannot be considered weak according to the criteria described in Staiger and Stock (1997).

2.5 Equilibrium Provision of Surplus Food

Next, we model the equilibrium provision of surplus food, or the number of products offered on the platform, and the price at which they are offered. Following Richards and Hamilton (2013) and Nair, Chintagunta, and Dube (2004), we assume the price of surplus food and the number of items offered on the platform are determined by solving the joint profit-maximization condition for the optimal amount and price for surplus harvest. Solving for the supply of surplus food products on the platform and the average price offered on the platform then allows us to derive hypotheses regarding the effect of shocks to platform demand on the price, supply of products offered, and profitability of the platform itself. As a structural model of surplus-food supply, the supply-side of our model allows to test the relative importance of platform-size and consumer traffic on the viability of our surplus-food CPMS.

At this time there are only a handful of firms in the surplus-harvest market in California. Still, pricing and output decisions are conditioned by the greater produce market as consumers have access to fresh produce that is closely substitutable for that sold by Imperfect Produce. Therefore, we assume platforms set category-level prices and assortments in a Bertrand-Nash manner for each category in the store. We derive the optimal retail pricing model first, followed by an expression for variety, or assortment depth.

Conditional on prices set by suppliers, the profit expression for the platform in time period t is written as:

$$\Pi_t = E[Q_t](p_t - r_t - w_t) - v(N_t), \tag{9}$$

where r_t is the constant cost of selling, w_t is an average wholesale price, $v(N_t)$ is the cost of expanding the size of the platform, and $E[Q_t]$ is the expected sales during week t, which is the product of the probability of household-purchase and the number of items purchased, aggregated to the market level. We follow Draganska and Jain (2005) in defining the cost of variety as a quadratic: $v(N_t) = \gamma_0 N_t + (1/2)\gamma_1 N_t^2$ ‘, which is appropriate as restocking costs can be expected to rise in a non-linear way with the number of products to be monitored, stored, re-shelved and priced. With this assumption, the platform's first-order condition in prices is given by:

$$E[Q_t] + \frac{\partial E[Q_t]}{\partial p_t}(p_t - r_t - w_t) = 0, \tag{10}$$

reflecting the local-monopoly assumption that the platform considers only the demand for their own products in setting prices. Stacking the first-order conditions for all items offered on the platform and time periods and solving for retail prices in matrix notation gives:

$$\mathbf{p} = \mathbf{r} + \mathbf{w} - \psi E[\mathbf{Q}]_p^{-1} E[\mathbf{Q}], \tag{11}$$

where $\mathbf{p}$ is a $JTx1$ vector of prices (J boxes, T weeks) , $\mathbf{w}$ is a $JTx1$ vector of wholesale prices, $\mathbf{r}$ is a $JTx1$ vector of product-specific selling-input prices, $E[\mathbf{Q}]$ is a $JTx1$ vector of expected quantities, and $E[\mathbf{Q}]_p$ is a diagonal $JTxJT$ matrix of expected-quantity-derivatives with respect to all retail prices. In the Imperfect Produce data, we do not observe the specific wholesale prices paid, so they are approximated by the indices of fruit and vegetable farm-gate prices described in more detail below. We parameterize these indices in the marginal cost equation as their true relationship with marginal cost is unknown, and must be estimated.

We estimate unobserved selling costs as a linear function of input prices, which is common practice in this literature (Villas-Boas 2007; Richards and Hamilton 2015). Specifically, we write the constant marginal cost of selling as: $r_{jt} = \delta_0 + \sum_{l=1}^{L} \delta_l v_l$,where $\mathbf{v}$ is a vector of input prices that includes a weekly measure of wages in the food retailing industry, an index of cardboard-box prices, a utility price index, and an index of energy costs. Finally, we parameterize the third term in equation (11) with a conduct parameter, ψ, to measure any departure from the maintained Bertrand-Nash pricing environment faced by Imperfect Produce. As is well understood in this literature, a value of $\psi = 1$ indicates that Imperfect Produce is able to price as a Bertrand-Nash oligopolist, but a value of $\psi = 0$ suggests that pricing approximates a perfectly competitive ideal.

Similarly, the first-order condition for optimal platform size is given by:

$$\frac{\partial E[Q_t]}{\partial N_t}(p_t - r_t - w_t) - \frac{\partial \nu}{\partial N_t} = 0, \tag{12}$$

where: $\partial\nu/\partial N_t$ is the marginal cost of augmenting the assortment, or increasing the size of the network. The first-order condition for platform size is also stacked over boxes and weeks to find:

$$\boldsymbol{\nu}_N = -E[\mathbf{Q}]_N E[\mathbf{Q}]_p^{-1} E[\mathbf{Q}], \tag{13}$$

where $\boldsymbol{\nu}_N = \gamma_0 + \gamma_1 \mathbf{N}$ is the marginal cost of adding to the network, and $E[\mathbf{Q}]_N$ is again a diagonal matrix of expected-quantity derivatives with respect to platform size. With this assumption, the estimated network-size equation is given by:

$$\mathbf{N} = -\tau_1 E[\mathbf{Q}]_N E[\mathbf{Q}]_p^{-1} E[\mathbf{Q}] - \tau_0, \tag{14}$$

where $\tau_0 = \gamma_0/\gamma_1$ and $\tau_1 = 1/\gamma_1$. Network size is a function of the equilibrium retail markup in (11), so the size of the network offered by the platform is function of wholesale prices. By estimating both margins and equilibrium network size together, we endogenize both decisions made by the platform manager, and thereby test the relative strength of network size on consumer demand for the platform, and the incentive to expand the number of products offered on it.

In the equilibrium-supply model derived in this section, we test the relative importance of consumer traffic and network size by estimating the equilibrium pricing and network-size equations together, and testing the importance of network size both directly and indirectly. Our direct test is a simple t-test of the sign and significance of the τ_1 parameter. In equilibrium, the marginal cost of adding another supplier must equal the marginal benefit. In equilibrium, if the value of τ_1 is positive, then adding another supplier provides positive incremental profit to the platform, all else constant. Because $\mathbf{p}$ and $\mathbf{N}$ are determined in equilibrium, however, we estimate the importance of indirect network effects by varying the level of $\mathbf{N}$ parametrically, and measuring the resulting impact on equilibrium margins by simulating our structural price- and network-model. Conditioned on the model of demand that drives both equilibrium pricing and margins, if a larger network causes margins to rise, then we have evidence of indirect network effects in the surplus-produce market.

On the supply side, the imputed-margin values, and marginal-value of network size are also likely to be endogenous. In order to instrument variables that reflect decisions taken by platform managers, we require variables that reflect demand-side shocks – shocks that are likely to be correlated with profitability and growth-potential, yet mean-independent of retail prices, and network-size, respectively. Given the limited data that we have available, the set of instruments for both equations are similar. Specifically, we capture box-specific demand shocks by including a vector of box-fixed effects. Second, we include our index of weekly fruit and vegetable shipments in order to capture any changes in demand that derive from the wider market for fruits and vegetables. Third, we capture any temporal changes in demand for the platform by including both a linear and quadratic time trend. Finally, we control for any remaining dynamic changes in demand by including lagged values of the imputed margin from equation (11), the marginal value of network size from equation (13), and lagged values of network size and retail price. For the pricing equation, a first-stage instrumental-variables regression produces an F-statistic of 15.03, and for the endogenous network size, these instruments produce an F-statistic of 15.16. In neither case can the instruments be described as weak (Staiger and Stock 1997). In the Results section below, we present estimates from both an IV and non-IV supply model in order to demonstrate the importance of controlling for endogeneity our system.

3 Data

Our data are from Imperfect Produce, Inc. Imperfect Produce began operations in mid-2015 with just over 1,000 customers, and by early 2017 had grown to over 7,500 customers. A cofounder of Imperfect Produce, Ron Clark, spent decades in the produce industry in California, and worked closely with the California Association of Food Banks. Through this association, he realized that there should be a market for surplus produce, or produce that does not make either formal grades set by either state marketing orders, or informal standards set by retail and foodservice buyers.[16]

Originally motivated by purely environmental and other social concerns, Imperfect Produce adopted, and adapted, a number of business practices over time in order to achieve financial independence. Imperfect Produce serves as an online, and real-life, platform that connects growers who have fresh fruit and vegetable items that either represent surplus harvest over contractual obligations or do not conform with usual size and quality standards in the grading process with retail or foodservice buyers looking for either low-cost meal ingredients or for "unloved" produce that they do not want to see go to waste. The key to the success of the Imperfect Produce platform is the immersive use of technology, from a state-of-the art mobile application to inventory-optimization software that rivals systems employed by larger, more established food distributors.

Similar to CPMS firms in other industries, Imperfect Produce recognized early on the value of data-capture, and data analytics in optimizing delivery schedules, and the mix of produce that they would want to source. Consequently, Imperfect Produce maintains a detailed database of every transaction they have ever executed, including all price, volume, and item specifications, as well as the amount of any promotion that was offered. Their data base provides a detailed description of the nature of each item, and how the customer assembled the box that they purchased. Box-prices for each transaction reflect a fixed amount for the box itself, and a variable amount for each item. After 2 years of operation, they have a detailed data set of every transaction, one that mimics the type of data collected by larger supermarkets, with millions more transactions. Imperfect Produce also captures all of their purchase-transactions, which we describe in more detail below.

Our data capture the entirety of transactions on the surplus food platform for the 60-week period from January 1, 2016 through February 28, 2017.[17] We observe the identity of the

[16] None of the marketing orders, or fruit and vegetable lobby groups, that we spoke with had any objection to the existence of firms selling off-grade produce because they represent ready markets for secondary-produce, create new markets for produce that would otherwise be thrown away, and enjoy overwhelming support from growers and packers. Implicitly, they support the price-discrimination role of secondary produce markets described in the Introduction.

[17] Our complete data set also contains transactions from September 1, 2015 - December 31, 2015, but IP

purchaser, the specific box that was purchased, the date it was purchased, the number and identity of the items in the box, the total amount spent, and the amount of any promotion used to capture the sale. Unfortunately, we have no demographic data on the purchasers to control for observed heterogeneity in preferences, although we do control for unobserved heterogeneity in the estimates presented below. All of our estimates are sufficiently precise that the lack of demographic data is not a great concern.

Table 1 summarizes the demand-side data. Perhaps most importantly, the summary data demonstrates considerable variation in the size of the Imperfect Produce platform, which varies from a low of 29 unique items to a high of 83 over the sample period. This suggests that variation in demand related to platform size is well-identified. Our summary data also show that customers demand a wide range of different boxes, and that order sizes can vary from small, single-box orders, to large orders in the hundreds of dollars. In terms of the attributes of each order, over 1/4 of all boxes are organic, and the vast majority consist of a mix of fruits and vegetables. Further, the plurality of boxes ordered are small, and the next most popular size is medium. Based on an average consumption rate of only 6.5 items per week (or roughly one per day), the popularity of small boxes is not surprising. That said, this consumption rate suggests that households using the Imperfect Produce platform are obtaining a substantial proportion of their fresh produce needs this way. In general, therefore, this summary data shows that there is likely sufficient variation in the data to identify the key parameters in our demand, and supply models, and that the Imperfect Platform plays an important role in users' shopping plans.

[table 1 in here]

We first use the IP data to determine whether there is summary evidence that the volumes of produce traded through the Imperfect Produce platform are related to the number of suppliers delivering produce to the site. Using the transactional data summarized in table 1, we show how much fresh produce has been sourced from growers over the first four years of

substantially changed their box-labeling system on January 1, 2016, so we removed the earlier transactions due to a lack of comparability across boxes.

operation, and how much has been transacted to buyers. Viability in two-sided networks is completely determined by demand from both sides of the market – from those who demand distributional services, and those who demand the end-product. In figure 1, we show that variation in user traffic appears to be positively related to changes in the number of suppliers to the platform, which is necessary for indirect network effects to arise. However, graphical evidence cannot control for many, potentially confounding factors.

[figure 1 in here]

We examine this question more closely by examining the data in figure 1 using a reduced-form approach in order to determine if sales volume is related to the number of items offered on the platform, and whether any relationship is robust to controlling for other factors that may explain the co-movement of traffic and supplier interest. Table 2 presents estimates from 3 models, each with slightly differing controls. In the first model (Model 1), the estimates show that price is the strongest determinant of sales volume, as expected, but the number of items offered on the platform, nonetheless, remains an important determinant of platform traffic. As in all of the reduced-form models, the attributes of each box are also important.

[table 2 in here]

In the next model (Model 2), we control for promotional spending because, as a startup, Imperfect Produce invested substantial amounts in building traffic. Controlling for spending does not alter the importance of price, and actually leads to a stronger role for platform size. Finally, we account for any temporal effects associated with platform demand in Model 3. Including a quadratic time-trend reduces the size of the price coefficient only slightly, and reduces the importance of network size by roughly 50%, but there appears to be no non-linear time-effects in the data. Because our data captures the latter part of their initial growth phase, controlling for a linear trend term is critical in the structural model below. In summary, our reduced-form evidence suggests that the number of suppliers, or platform size, appears to be critical to growing demand for a surplus-produce platform.

We also have access to all of the firm's purchase data, albeit in lesser detail than the transaction data described above. While we have records of individual purchases, so we can track what was purchased and whom it was purchased from, we do not have volume data specific enough to assign a cost to each individual item. Nonetheless, we have total purchasing costs by week and, given that Imperfect Produce does not hold substantial inventories of any item, we are able to impute a rough estimate of their gross margin over time. While volume is a necessary condition for success, it is not sufficient without a margin-generating mechanism.

Some platforms in the surplus harvest business serve as brokers, extracting a fee on each transaction negotiated between a farmer with surplus produce, and a retail or foodservice buyer. Imperfect Produce, on the other hand, conducts business as a traditional middleman, taking title to the produce purchased from growers, and selling for their own account to individual customers. In figure 2, we show how Imperfect Produce's gross margin has varied over time, by month, and how the margin has varied with the number of items offered on the site. Interestingly, there appears to be a strong, positive relationship between gross margin and the number of items offered on the site over the first 12 months of our sample period (calendar 2016) (linear regression coefficient $= 3,299$, t-ratio $= 2.352$), the relationship all but disappears over the final 2 months of the sample (linear regression coefficient $= 324$, t-ratio $= 0.291$). Although January and February are relatively slow months in the fresh produce business, Imperfect Produce management assures us that these months were not out of the ordinary.

[figure 2 in here]

Data for our instruments and selling costs are from the USDA, Agricultural Marketing Service and the Bureau of Labor Statistics. Specifically, the fresh produce price indices and shipment values are from the USDA, on a monthly basis. We do nos adjust the monthly frequency to smooth out variation to estimate with the other, weekly data series.[18] Our input price series are from the Bureau of Labor Statistics, Producer Price Index series, again

[18] While doing so is common, we believe fitting a cubic-spline, or similar smoothing method, imputes artificial variation in the data that is likely to be misleading.

on a monthly basis. We capture labor costs, which are the largest component of selling costs, by including a measure of weekly compensation to workers in the food retailing industry. Another large, and unique, cost item for Imperfect Produce is the price of boxes used to deliver items to retail customers. We measure the cost of packaging with a monthly index of prices earned by cardboard-box manufacturers. Utility and fuel costs are also producer-price indices for generators and distributors of electricity, and fuel wholesalers, respectively. Although Imperfect Produce faces a number of other costs of doing business, our price indices are likely to capture most of the variation in the marginal cost of buying, handling, and distributing surplus produce.

4 Results and Discussion

We first present our estimates of the demand-side model, followed by estimates of the pricing-and-network size model, and then demonstrate the importance of our findings for food policy through numerical simulation. Within each stage, we present estimates from a range of specifications to evaluate the robustness of our model and conduct a series of specification tests to determine the preferred model.

4.1 Platform Demand Estimates

In this section, we begin by presenting the estimates from each of a series of platform demand models, and then interpret our findings from the best-fitting specification of the model. Table 3 shows our platform demand estimates. Our initial specification, which is not shown in the table, maintains a simple Logit-Poisson process; however, tests for overdispersion revealed that the Negative Binomial was preferred ($\chi^2 = 640.7$). Within the class of Negative Binomial models, moreover, the Negative-Binomial-P (NBP) model, which is the most general specification we estimate, is the preferred model. Consequently, the entries in Table 3 present the results of three different versions of the Logit-NBP model.

[table 3 in here]

Model 1 in Table 3 makes no attempt to control for endogeneity of either the platform price or the size of the network (number of items offered). Identifying network effects when both variables are endogenous is the focus of much of the recent empirical network economics literature (Jeziorski 2014, for example), effects of particular importance in our setting given the intent of the purchase incidence model to capture need-based motivations for visiting the Imperfect Produce site. Because the incidence and purchase quantity equations are estimated together, we include controls for endogenous prices and network size in Model 2 to remove any bias that may be transmitted to the incidence equation. Comparing the estimates in Model 2 to those in Model 1, we see that Model 2 is preferred ($\chi^2 = 21,668$, based on a likelihood-ratio test) and removes a substantial amount of bias from the key parameters in the incidence model, σ and θ.

Model 3 accounts for unobserved heterogeneity by allowing for random household-level parameters. Comparing the fit of Model 2 to Model 3 reveals Model 2 to be the preferred specification. Although the scale parameters on the σ and θ are statistically significant, they are very small in magnitude, which indicates that heterogeneity is not an important consideration. For this reason, we use the estimates from Model 2 to test our hypotheses regarding the importance of network-size on demand, and to condition our supply-model estimates to follow.

Based on the Purchase Incidence estimates in Table 3, it is apparent that consumption rate, inter-purchase time, and the amount purchased on the previous visit are all critically important in the binary decision to order surplus food from the platform. This result is not surprising in light of the prior empirical literature on purchase incidence (Briesch, Chintagunta, and Fox 2009). Conditioned on these variables, we also find that promotions are not important drivers of site traffic, but that the attributes of the order are important. Customers prefer organic boxes, either in a small or medium size, consisting of a mix of fruits and vegetables.

With respect to the Purchase Quantity component, the estimates in the bottom panel of Table 3 are statistically significant in determining the amount purchased in each order. The estimated value for the Network Size variable implies a net marginal effect of 0.05, meaning that an increase of one item in the assortment on offer will result in an expected increase of 0.05 items per week, accounting for the equilibrium effect on purchase incidence, and the number of items purchased if an order is placed. In elasticity terms, this estimate suggests a network-size elasticity of 0.23, which implies that a 10% increase in network size can be expected to result in a 2.3% increase in expected sales. The price elasticity implied by the estimates in Table 3 is -0.15, which indicates that the quantity response to changes in surplus food prices is highly inelastic. This is perhaps to be expected given that the clientele for Imperfect Produce may order surplus food not only for the purpose of obtaining inexpensive produce, but also to contribute to resolving the problem of food waste. Unlike in the incidence model, we find that promotion has an important impact on the quantity purchased, and that users tend to order significantly more surplus food items when purchasing organic relative to conventional produce.

As in the incidence model, we find that the controls for the endogeneity of price and network size are statistically significant. The importance of controlling for endogeneity is emphasized by comparing the estimates from Model 2 to those of Model 1. While there appears to be little bias in estimated price effect, the marginal effect of network size is estimated with substantial bias in the non-control function model. These demand estimates, in turn, drive the equilibrium supply model estimates described next.

4.2 Pricing and Network Size

Conditioned on consumer preferences for platform size and prices, the supply estimates test for the effect of network size and equilibrium pricing. In this model, we control for the simultaneous influences of pricing and demand on the marginal value of increasing the size of the platform. Although our demand estimates above show that consumers have a positive marginal value of network-size, the converse must also be true to create a "virtuous cycle" of indirect network effects: Namely, suppliers must also value consumer traffic and be more likely to use the platform as the size of the network increases. In our model, we infer the demand for platform services by estimating the marginal value of network-size demand by the platform manager. In Table 4, this estimate is shown by the value of τ_1, which is the equilibrium marginal value of N from the manager's perspective.

Notice that our estimates of τ_1 in Table 4 are remarkably robust across the three specifications we consider. In Model 1, we estimate τ_1 independent from the equilibrium-pricing equation, while in Model 2 we estimate both equations together, but without endogeneity controls. In Model 3, we estimate both equations together, and with endogeneity controls. Because the estimators are different for each model, we compare goodness-of-fit using a simple non-nested comparitor, the pseudoR^2. By this measure, we see that Model 3 provides the best fit to the data, while also controlling for the endogeneity of both marginal network value, and retail margins.

Model 3 reveals the marginal network value estimate to be positive and significant, as in the other models, but with a magnitude that is more than twofold the size. Both the margin and network-value models in Table 4 are estimated in money-metric terms, which allows us to interpret this estimate as the marginal value to the network operator, per box, of an additional item in the assortment, or \$0.55 / item. Based on an average box price of \$6.47 (box size times item price in table 1), this estimate represents fully 8.5% of the retail value of a box. More importantly, it suggests that, in equilibrium, the platform manager is willing to pay \$0.55 / item on a per-box basis, for another item in the assortment. Because this is an estimate of the equilibrium value, it suggests that there is a significant, positive value to distribution through the platform from produce suppliers. Put differently, our estimate implies that suppliers are willing to accept \$0.55 / item for distribution, indicating that when consumers demand more items on the platform suppliers demand greater distribution through the platform.

[table 4 in here]

Estimates from the pricing model in Table 4 show how properly accounting for the endogeneity of network size can have dramatic effects on the degree of market power exercised by the platform manager. While estimating either the pricing- and network-size equations separately (Model 1) or without endogeneity controls (Model 2), the outcome implies nearly-competitive pricing conduct (ψ is near zero in each case); however, when the marginal value of N and retail margins are properly instrumented, the estimate of ψ is closer to Bertrand-Nash behavior than to competitive pricing. Consequently, this finding suggests that platform managers in CPMS markets such as this are able to exploit indirect network effects to generate much higher margins than would otherwise be the case.

4.3 Policy Implications

In this Section we numerically characterize the policy implications of subsidizing CPMS markets to reduce the amount of produce that is not purchased for human consumption. Our numerical model illustrates the importance of platform size, the demand for distribution services by providers, the demand for surplus food by consumers, and platform pricing in determining surplus food transactions on the CPMS platform. We allow for endogenous consumer demand, retail prices (and margins), and network-size in our numerical model, which allows us to demonstrate the importance of indirect network effects in the market by varying parameters in consumer demand and by introducing our key food policy instruments. We first examine the effect of changes in consumer acceptance for surplus food.

Table 5 presents estimates of the retail margin, and platform-demand for network-size by varying consumer demand for surplus food. Specifically, we vary the key parameter in the Order-Size demand model, ϕ_N in Table 3, that measures consumers' preferences for purchasing surplus food on the CPMS platform. Increasing consumers' preference for the number of items available on the platform should lead to three effects: (i) higher prices; (ii) an increase in the number of surplus food items offered on the platform; and (iii) indirectly through prices, reduced quantity demanded for each surplus food product on the platform.[19] Our numerical simulation precisely reveals these effects. Higher values of ϕ_N are indeed associated with significantly higher retail prices (and margins) and a significant increase in network size. The reason is that a positive shock in consumers' preferences for product variety causes vendors to respond by placing a greater number of products on the platform, increasing demand for each item on the platform, resulting in higher equilibrium prices. These are the necessary ingredients for producing indirect network effects: Platform managers are able to leverage greater demand from one side of the market (consumers) to increase demand for distribution from the other side of the market (suppliers) in a virtuous cycle that raises platform rents.

[table 5 in here]

We are now ready to examine the policy implications of our findings for surplus food. Specifically, we consider a price-based incentive, which we refer to as an "ugly food" subsidy, that provides a price incentive for consumers to purchase surplus food on the CPMS platform. Table 6 shows the effect of an ad valorem subsidy on produce purchased on the platform, where we define the net price paid by consumers as $\widehat{p} = (1 - \eta)p$, p as the list price and η as the subsidy rate.

Given the non-linearity of the structural model, the effect of varying the subsidy from 10% through 90% results in a highly non-linear response for both the equilibrium price and the number of suppliers. A relatively small subsidy level (25%) produces only modest changes in the overall size of the network (48.5 versus 46.9 products on offer), while the effect on price is more substantial ($1.72 / item versus $1.57 / item). The total quantity of surplus food purchased on the platform rises from 17.42 to 27.92 items per order – a 60% increase that siphons surplus food products onto the platform that would otherwise result in food loss.[20]

[19] We assume that quality does not decline in response to a demand shock. When demand rises, managers at the firm reach out to more suppliers, and source more volume from existing suppliers. Still, they offer produce with only cosmetic imperfections, and never nutritional or eating deficiencies.

[20] Surplus food sold through Imperfect Produce may have otherwise been used for animal feed, composted, or used in some other, low-value purpose. However, we cannot make a general statement as to alternative destinations for the food sold through the platform. Ben Chesler, CEO of Imperfect Produce, states that "...It is very specific to the item and specific purchase. Generally it goes to waste but we avoid making broad claims like that. Some goes to waste, some foes to animal feed, some goes to processors..." More generally,

At higher subsidy levels (90%), the effect on network size increases sharply (75.33 versus 46.9 products on offer), while the price effect is moderate ($2.11 / item versus $1.57 / item). The increase in the quantity of surplus food sold on the platform is substantial, more than tripling the quantity of surplus food purchased on the platform (from 17.42 to 53.69). Policies that provide monetary incentives to consumers for purchasing surplus food products have a dramatic impact on the amount of food lost to the retail market.

[table 6 in here]

5 Conclusions

In this paper, we study the viability of a secondary market in the sharing economy for ugly produce, or fresh fruits and vegetables that fail to make more usual saleable-grades. We develop a model of a two-sided market in which consumers demand a variety of produce from a surplus-harvest website, and suppliers seek distribution to the greatest number of consumers possible. We estimate our empirical model using transactional data from a sharing-economy firm that has been operating successfully in this area for two years, Imperfect Produce, LLC.

Our results show that, controlling for prices and the attributes of the items they are purchasing, consumer demand for deliveries through the Imperfect Produce website rise in the variety of items they offer. Controlling for the endogeneity of surplus produce-supplies, we show that equilibrium margins and distribution rise in the number of items offered on the site, which supports our hypothesis that Imperfect Produce operates in a two-sided market. As is the case with other two-sided markets, the platform manager is able to use indirect network effects to his or her advantage, increasing user-demand for the website by providing more items, and generating demand for distribution by attracting a greater number

our assumption on this point is that the farmer allocates surplus harvest to its highest and best use. By creating an active market for produce, in direct-consumption form, the value of food sold through IP is likely to be (we do not have prices for alternative uses such as feed or processing) much greater. While this is not reducing waste in the sense of Bellemare et al. (2017), we agree with Buzby et al. (2011) who argue more generally that food losses include the degradation in value of food such that it cannot be used for its intended purpose. That is, because producers purchase scarce inputs such that their price is equal to their marginal value product, and marginal value product depends on the output price, degradation in saleable value represents a misuse of inputs, and economic loss.

of consumers. Counter-factual simulations of our equilibrium pricing and network-size model show that increasing the intensity of demand for variety, or the number of items offered on the site, increases the demand for distribution by suppliers by an elasticity of roughly 0.5. We interpret this result as demonstrating that sharing-economy firms in the surplus-food market can take advantage of substantial indirect network effects.

Finding evidence of indirect network effects in a non-conventional, sharing-economy industry represents a substantial contribution to both the empirical literature on how CPMS firms operate, and the practical literature on food waste. Ours is the first study to document indirect network effects in CPMS firms, which should be a necessary condition for their success. While the viability of any startup firm is not guaranteed, if the fundamental economics of the industry suggest the presence of indirect network effects, then success is substantially more likely.[21] There is plenty of evidence of indirect network effects in technology industries, from personal digital assistants (PDAs, Nair, Chintagunta and Dube 2004) to yellow pages (Rysman 2004), but we are the first to show they also exist in the sharing economy.

In terms of the surplus food problem, our findings show that there may indeed be a market solution to an issue that has otherwise been regarded as largely intractable, resulting from behavioral errors by millions of agents in the economy, each with limited ability to solve the errors-in-planning that result in either surplus harvest, or food that perishes before it can be used. If the conditions exist for a market to arise in surplus food, then at least farmers will have an incentive to manage their harvests optimally. In fact, platforms such as Food Cowboy have emerged to provide households and restaurants a means of selling food they would otherwise throw away, providing a first step in making a market for food waste more generally. While the larger waste problem is not likely to disappear, economists understand

[21] Unlike many technology platforms that emerge as monopoly, or near–mononoply, due to the fact that they are subject to indirect network effects, and their platform is not compatible with others (Apple iPhone, and Microsoft Windows, for example), consumers in the surplus harvest market have the ability to multi-home (Armstrong 2006) so the same competitive lock-in that we see in technology markets is not likely to occur. We define success in this market, therefore, as being able to generate sufficient long-term profit to survive, and not to earn monopoly profits as in other industries that are also subject to indirect network effects.

that aligning incentives with the larger social objective – minimizing food waste – can move us toward a longer-term solution to the problem.

Our findings have important implications for both the management of sharing-economy platforms like Imperfect Produce, and the viability of surplus-produce trading more generally. In any sharing-economy platform, the manager intermediates between users who demand a greater breadth of service and suppliers who agree to share their surplus goods with potential users. In the case of Imperfect Produce, users of the food surplus website are attracted by the variety of items on offer, and suppliers are attracted by the number of consumers on the site. If the fundamental economics of two-sided markets continues to work as we have shown here, then greater expansion of the concept beyond surplus harvest to leftover perishables from retail stores, household compost and restaurant-waste are indeed possibilities.

Our research has some important limitations. Most importantly, the Imperfect Produce data describes the operation of a startup firm, in an industry that is struggling to become established. To the extent that management was learning-on-the-job while our data were being generated, it likely contains more noise than would be the case if the data were generated by a more established firm. Second, Imperfect Produce did not record accurate, per-item data from suppliers. With more accurate data on the demand-for-distribution from suppliers, we would have been able to estimate the demand for distribution in a more direct way. Third, and perhaps most importantly, the market for surplus harvest is currently so small that it would be a heroic effort to attempt to infer any aggregate welfare effects due to this platform. However, Chen, Esteban, and Shum (2013) show that such markets for secondary output may, in fact, harm the interests of the firms involved. We leave this question for future research. Finally, our data describe only the California market. Whether they will generalize to the larger US, or global, markets, is uncertain.

6 Appendix: The NegBin-P Model

In this appendix, we derive the alternative count-data model to our maintained Poisson specification. This specification, the Negative-Binomial-P (NBP) model (Greene 2010) accounts for overdispersion typical in count-data settings, in a very general way. Given that the usual form of the NegBin model is given as (Greene 2003):

$$\Pr(Y = y_i|\mathbf{x}_i) = \frac{\Gamma(1/\alpha + y_i)}{\Gamma(1/\alpha)\Gamma(y_i + 1)} u_i^{1/\alpha}(1 - u_i)^{y_i}, \tag{15}$$

where $u_i = 1/(1 + \alpha\lambda_i)$, then we replace $1/\alpha$ with $(1/\alpha)\lambda_i^{2-P}$ such that:

$$\Pr(Y = y_i|\mathbf{x}_i) = \left(\frac{\Gamma((1/\alpha)\lambda_i^Q + y_i)}{\Gamma((1/\alpha)\lambda_i^Q)\Gamma(y_i + 1)}\right)\left(\frac{(1/\alpha)\lambda_i^Q}{(1/\alpha)\lambda_i^Q + \lambda_i}\right)^{(1/\alpha)\lambda_i^Q}\left(\frac{\lambda_i}{(1/\alpha)\lambda_i^Q + \lambda_i}\right)^{y_i}, \tag{16}$$

where $Q = 2 - P$. In the combined Logit-Poisson platform-demand model, we substitute the NBP specification for the Poisson stage in order to test for the importance of overdispersion in estimating the strength of indirect network effects in the Imperfect Produce platform. We compare the two specifications using the simple OLS-based test described in the text.

References

[1] Ackerberg, D. A. & Gowrisankaran, G. (2006). Quantifying equilibrium network externalities in the ACH banking industry. *RAND Journal of Economics*, 37(3), 738-761.

[2] Acuff, K., & Kaffine, D. T. (2013). Greenhouse gas emissions, waste and recycling policy. *Journal of Environmental Economics and Management*, 65(1), 74-86.

[3] Alexander, C., & Smaje, C. (2008). Surplus retail food redistribution: An analysis of a third sector model. *Resources, Conservation and Recycling*, 52(11), 1290-1298.

[4] Argentesi, E., & Filistrucchi, L. (2007). Estimating market power in a two-sided market: the case of newspapers. *Journal of Applied Econometrics*, 22(7), 1247-1266.

[5] Armstrong, M. (2006). Competition in two-sided markets. *RAND Journal of Economics*, 37, 668-681.

[6] Bardhi, F., & Eckhardt, G. M. (2012). Access-based consumption: The case of car sharing. *Journal of Consumer Research*, 39(4), 881-898.

[7] Belk R. (2014). You are what you can access: sharing and collaborative consumption online. *Journal of Business Research* 67(8), 1595–1600.

[8] Bell, D. R., Ho, T. H., & Tang, C. S. (1998). Determining where to shop: Fixed and variable costs of shopping. *Journal of Marketing Research*, 35(3), 352-369.

[9] Bellemare, M. F., Çakir, M., Peterson, H. H., Novak, L., & Rudi, J. (2017). On the Measurement of Food Waste. *American Journal of Agricultural Economics*. forthcoming.

[10] Botsman, R. & Rogers, R. (2010). *What's mine is yours: the rise of collaborative consumption*. New York: HarperCollins.

[11] Botsman, R. (2013). The sharing economy lacks a shared definition. (https://www.fastcompany.com/3022028/the-sharing-economy-lacks-a-shared-definition). Accessed on Dec. 11, 2017.

[12] Briesch, R. A., Chintagunta, P. K., & Fox, E. J. (2009). How does assortment affect grocery store choice? *Journal of Marketing Research*, 46(2), 176-189.

[13] Bucklin, R. E., Gupta, S., & Siddarth, S. (1998). Determining segmentation in sales response across consumer purchase behaviors. *Journal of Marketing Research*, 35(2), 189-197.

[14] Buzby, J. C., Hyman, J., Stewart, H., & Wells, H. F. (2011). The value of retail-and consumer-level fruit and vegetable losses in the United States. *Journal of Consumer Affairs*, 45(3), 492-515.

[15] Buzby, J. C., & Hyman, J. (2012). Total and per capita value of food loss in the United States. *Food Policy*, 37(5), 561-570.

[16] Buzby, J. C., Farah-Wells, H., & Hyman, J. (2014). The estimated amount, value, and calories of postharvest food losses at the retail and consumer levels in the United States. *USDA-ERS Economic Information Bulletin*, (121).

[17] Caillaud, B., & Jullien, B. (2003). Chicken & egg: Competition among intermediation service providers. *RAND Journal of Economics*, 34(2): 309-328.

[18] Cameron, A. C., & Trivedi, P. K. (1990). Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics*, 46(3), 347-364.

[19] Caswell, J. (1998). How labeling of safety and process attributes affects markets for food. *Agricultural and Resource Economics Review*, 27(2), 151-158.

[20] Chandra, A., Collard-Wexler, A. (2009). Mergers in two-sided markets: an application to the Canadian newspaper industry. *Journal of Economics & Management Strategy*, 18, 1045-1070.

[21] Chen, J., Esteban, S., & Shum, M. (2013). When do secondary markets harm firms?. *American Economic Review*, 103(7), 2911-2934.

[22] Chu, J. & Manchanda, P. (2016). Quantifying cross and direct network effects in online C2C platforms. *Marketing Science*, 35(6), 870-893.

[23] Clements, M. T., & Ohashi, H. 2005. Indirect network effects and the product cycle: Video games in the US, 1994-2002. *Journal of Industrial Economics* 53 (4): 515-

[24] Corts, K. S., & Lederman, M. 2009. Software exclusivity and the scope of indirect network effects in the US home video game market. *International Journal of Industrial Organization* 27(2): 121-

[25] Coleman-Jensen, A., Gregory, C., & Singh, A. (2014). Household food security in the United States in 2013. *USDA-ERS Economic Research Report*, (173).

[26] Dinan, T. M. (1993). Economic efficiency effects of alternative policies for reducing waste disposal. *Journal of Environmental Economics and Management*, 25(3), 242-256.

[27] Dixit, A. K., & Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *American Economic Review*, 67(3), 297-308.

[28] Dubé, J.- P. H., G. J. Hitsch, & Chintagunta, P. K. 2010. Tipping and concentration in markets with indirect network effects. *Marketing Science* 29(2): 216-229.

[29] Fraiberger, S. P., & Sundararajan, A. (2015). Peer-to-peer rental markets in the sharing economy. NYU Stern School of Business Research Paper.

[30] Fullerton, D., & Kinnamanm T. C. (1995). Garbage, recycling, and illicit burning or dumping. *Journal of Environmental Economics and Management*, 29(1), 78-91.

[31] Fullerton, D., & Wu, W. (1998). Policies for green design. *Journal of Environmental Economics and Management*, 36(2), 131-148.

[32] Gandal, N., M. Kende, & Rob, R. (2000). The dynamics of technological adoption in hardware / software systems: the case of compact disc players, *RAND Journal of Economics*, 31(1), 43-61.

[33] Garfield, L. (2016). This new 'dating app' for ugly fruits and veggies has already saved 20,000 pounds of produce. *TechInsider*. (http://www.techinsider.io /cerplus-is-like -a-dating-app-for- ugly-produce-2016-6). Accessed July 21, 2016.

[34] Garrone, P., Melacini, M., & Perego, A. (2014). Opening the black box of food waste reduction. *Food Policy*, 46, 129-139.

[35] Greene, W. H. (2003). *Econometric Analysis* (5th). Ed.. Upper Saddle River, NJ.

[36] Gustavsson, J., Cederberg, C., Sonesson, U., Van Otterdijk, R., & Meybeck, A. (2011). *Global food losses and food waste*. Food and Agriculture Organization of the United Nations, Rome.

[37] Hall, K. D., Guo, J., Dore, M., & Chow, C. C. (2009). The progressive increase of food waste in America and its environmental impact. *PloS one*, 4(11), e7940.

[38] Halloran, A., Clement, J., Kornum, N., Bucatariu, C., & Magid, J. (2014). Addressing food waste reduction in Denmark. *Food Policy*, 49, 294-301.

[39] Hodges, R. J., Buzby, J. C., & Bennett, B. (2011). Postharvest losses and waste in developed and less developed countries: opportunities to improve resource use. *The Journal of Agricultural Science*, 149(S1), 37-45.

[40] Jeziorski, P. (2014). Effects of mergers in two-sided markets: The US radio industry. *American Economic Journal: Microeconomics*, 6(4), 35-73.

[41] Jin, G. Z., & Rysman, M. (2015). Platform pricing at sports card conventions. *Journal of Industrial Economics*, 63(4), 704-735.

[42] Kaiser, U. and J. Wright. (2006). Price structure in two-sided markets: evidence from the magazine industry. *International Journal of Industrial Organization*, 24, 1-28.

[43] Kantor, L. S., Lipton, K., Manchester, A., & Oliveira, V. (1997). Estimating and addressing America's food losses. *Food review*, 20(1), 2-12.

[44] Kim, J., Allenby, G. M., & Rossi, P. E. (2002). Modeling consumer demand for variety. *Marketing Science*, 21(3), 229-250.

[45] Lamberton C. P., & Rose, R. L. (2012). When is ours better than mine? A framework for understanding and altering participation in commercial sharing systems. *Journal of Marketing* 76: 109–125.

[46] Lee, R. S. (2013). Vertical integration and exclusivity in platform and two-sided markets. *American Economic Review*, 103(7), 2960-3000.

[47] McAlister, L., & Pessemier, E. (1982). Variety seeking behavior: An interdisciplinary review. *Journal of Consumer Research*, 9(3), 311-322.

[48] Möhlmann, M. (2015). Collaborative consumption: determinants of satisfaction and the likelihood of using a sharing economy option again. *Journal of Consumer Behaviour*, 14(3), 193-207.

[49] Mussa, M. and S. Rosen. (1978). Monopoly and product quality. *Journal of Economic Theory,* 18(2), 301-317.

[50] Nair, H., P. Chintagunta, and J.-P. Dube. (2004). Empirical analysis of indirect network effects in the market for personal digital assistants. *Quantitative Marketing and Economics,* 2, 23-58.

[51] Parfitt, J., Barthel, M., & Macnaughton, S. (2010). Food waste within food supply chains: quantification and potential for change to 2050. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1554), 3065-3081.

[52] Park, S. (2004). Quantitative analysis of network externalities in competing technologies: The VCR case. *Review of Economics and Statistics* 86(4): 937-945.

[53] Petrin, A., & Train, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research*, 47(1), 3-13.

[54] Richards, T. J., & Hamilton, S. F. (2013). Network externalities in supermarket retailing. *European Review of Agricultural Economics*, 40(1), 1-22.

[55] Richards, T. J., & Hamilton, S. F. (2015). Variety pass-through: An examination of the ready-to-eat breakfast cereal market. *Review of Economics and Statistics*, 97(1), 166-180.

[56] Rochet, J. C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association*, 1(4), 990-1029.

[57] Rochet, J. C., & Tirole, J. (2006). Two-sided markets: a progress report. *RAND Journal of Economics*, 37(3), 645-667.

[58] Rysman, M. (2004). Competition between networks: a study of the market for yellow pages. *Review of Economic Studies,* 71, 483-512.

[59] Steiner, M., Wiegand, N., Eggert, A., & Backhaus, K. (2016). Platform adoption in system markets: the roles of preference heterogeneity and consumer expectations. *International Journal of Research in Marketing*, 33(2), 276-296.

[60] Sundararajan, A. (2013). From Zipcar to the sharing economy. *Harvard Business Review*, 1.

[61] Sundararajan, A. (2014). Peer-to-peer businesses and the sharing (collaborative) economy: Overview, economic effects and regulatory issues. Written testimony for the hearing titled *The Power of Connection: Peer to Peer Businesses*, January.

[62] Theotokis, A., Pramatari, K., & Tsiros, M. (2012). Effects of expiration date-based pricing on brand image perceptions. *Journal of Retailing*, 88(1), 72-87.

[63] Tsiros, M., & Heilman, C. M. (2005). The effect of expiration dates and perceived risk on purchasing behavior in grocery store perishable categories. *Journal of Marketing*, 69(2), 114-129.

[64] United Nations. (2015). World population projected to reach 9.7 billion by 2050. (http://www.un.org/en/development/desa/news/population/2015-report.html). Accessed on July 21, 2016.

[65] United States Environmental Protection Agency. (2016). Turning food waste into energy at the East Bay Municipal Utility District (EBMUD). (https://www3.epa.gov /region09/ waste/features/ foodtoenergy/food-waste.html). Accessed on July 20, 2016.

[66] Van Cayseele, P. J. G., and S. Vanormelingen (2009). Prices and network effects in two-sided markets: the Belgian newspaper industry. Working paper, University of Leuven, Leuven, Belgium.

[67] Zhou, Y. (2016). Bayesian estimation of a dynamic model of two-sided markets: application to the U.S. video game industry. *Management Science*, Articles in Advance, pp. 1–21.

Table 1. Data Summary

Variable	Units	Mean	Std. Dev.	Min.	Max.	N
Number of Items	#	46.3801	14.5647	29	83	201836
Box Size	#	7.5409	3.9272	4	12	201836
Order Dollars	$	20.3375	11.1023	6	450	201836
Order Items	#	13.5346	7.6526	4	381	201836
Promotion Dollars	$	0.2719	2.3352	0	140.12	201836
Item Price	$ / Item	1.5710	0.2727	0.38	11.18	201836
Organic	%	26.4185	44.0899	0	100	201836
Fruit	%	1.9283	13.7518	0	100	201836
Vegetable	%	2.2389	14.7947	0	100	201836
Small	%	24.5338	43.0288	0	100	201836
Medium	%	21.2296	40.8934	0	100	201836
Large	%	4.4452	20.6097	0	100	201836
Consumption Rate	Items / Week	6.5507	3.5494	1.19	72.46	201836
Box 1	%	0.9840	9.8706	0	100	201836
Box 2	%	0.9443	9.6717	0	100	201836
Box 3	%	0.6733	8.1779	0	100	201836
Box 4	%	1.7262	13.0245	0	100	201836
Box 5	%	8.9355	28.5256	0	100	201836
Box 6	%	15.0835	35.7889	0	100	201836
Box 7	%	0.7030	8.3553	0	100	201836
Box 8	%	2.7190	16.2638	0	100	201836
Box 9	%	10.2608	30.3447	0	100	201836
Box 10	%	7.3163	26.0405	0	100	201836
Box 11	%	1.0494	10.1900	0	100	201836
Box 11	%	1.1896	10.8417	0	100	201836

Note: Data from Imperfect Produce, LLC.

Table 2. Reduced-Form Sales Volume Regression

	Model 1		Model 2		Model 3	
	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
Constant	34.3496*	0.1444	33.1989*	0.1393	32.6954*	0.1605
Price	-7.2329*	0.0684	-7.2331*	0.0657	-7.0529*	0.0691
Number of Items	0.5574*	0.1045	0.9407*	0.1005	0.4071*	0.1552
Organic	3.8004*	0.0334	3.6792*	0.0321	3.6016*	0.0330
Fruit	0.2259*	0.0738	0.1935*	0.0709	0.1797*	0.0709
Veg	0.0445	0.0691	0.0608	0.0664	0.0519	0.0664
Small	-15.0453*	0.0870	-14.1162*	0.0842	-14.1456*	0.0843
Medium	-12.0483*	0.0860	-11.2289*	0.0832	-11.2091*	0.0831
Large	-6.7192*	0.0952	-6.0758*	0.0918	-6.0557*	0.0917
Promotion			0.3793*	0.0041	0.3807*	0.0041
Week					0.0185*	0.0043
Week^2					-0.1105*	0.0726
R^2	0.4582		0.4997		0.5002	
F	11,006.35		11,554.05		9,472.21	

Note: A single asterisk indicates significance at a 5% level. Data are at the household level, on a weekly basis.

Table 3. Demand Estimates: Logit / NB-P Model

	Model 1		Model 2		Model 3	
	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
Logit Purchase Incidence Model						
σ	0.7736*	0.0905	0.7909*	0.1197	0.7790*	0.0938
σ(s)					0.0100*	0.0023
θ	0.7342*	0.2752	0.7406	0.4646	0.7362*	0.1597
θ(s)					0.0094	0.0064
Consumption Rate	30.6115*	0.1214	30.6144*	0.1221	30.6124*	0.1195
Inter. Time	21.2850*	0.1263	21.2863*	0.1270	21.2854*	0.1176
Lagged Q	-4.1132*	0.0550	-4.1098*	0.0555	-4.1121*	0.0438
Promotion	-0.2807*	0.0020	-0.2122*	0.0019	-0.2583*	0.0016
Week	-6.9485*	0.0348	-6.9299*	0.0421	-6.9425*	0.0357
Organic	0.1746*	0.0786	0.1925*	0.0998	0.1800*	0.0728
Fruit	-6.3739*	0.1073	-6.3655*	0.1332	-6.3712*	0.1036
Vegetable	-4.5805*	0.1733	-4.5717*	0.2000	-4.5777*	0.1386
Small	7.9059*	0.0974	7.9236*	0.1420	7.9113*	0.0953
Medium	6.4082*	0.1002	6.4257*	0.1467	6.4138*	0.1032
Large	3.0118*	0.1033	3.0207*	0.1504	3.0146*	0.0610
Price Control			0.5004*	0.0399	2.7108*	0.0306
Network Control			0.4998*	0.0851	7.2893*	0.0745
NB-P Purchase Quantity Model						
Constant	3.9838*	0.0018	3.9856*	0.0024	3.9852*	0.0001
Price	-0.7018*	0.0006	-0.7103*	0.0012	-0.6997*	0.0008
Network Size	0.0155*	0.0002	0.0150*	0.0002	0.0040*	0.0000
Promotion	0.0286*	0.0002	0.0292*	0.0002	0.0332*	0.0000
Organic	0.3316*	0.0007	0.3189*	0.0013	0.3307*	0.0003
Fruit	-0.0040	0.0044	-0.0030	0.0042	-0.0039	0.0043
Vegetable	-0.0170*	0.0070	-0.0165*	0.0065	-0.0170*	0.0067
Small	-0.8517*	0.0007	-0.8613*	0.0022	-0.8520*	0.0003
Medium	-0.5828*	0.0020	-0.5758*	0.0019	-0.5820*	0.0011
Large	-0.2569*	0.0005	-0.2537*	0.0015	-0.2565*	0.0003
Lambda(s)					0.0351*	0.0000
Price Control			0.4886*	0.0003	-0.0149*	0.0000
Network Control			0.4986*	0.0011	0.0099*	0.0001
T	0.0054*	0.0008	0.0059*	0.0003	0.0455*	0.0000
Q	6.4870*	0.0059	6.4870*	0.0088	6.4870*	0.0001
LLF	-540739		-558033		-542233	
AIC	5.6130		5.793		5.629	

Note: Model 1 = Logit / NB-P model without controls, fixed parameters.

Model 2 = Logit / NB-P model with controls, fixed parameters.

Model 3 = Logit / NB-P model with controls, random parameters

Table 4. Pricing and Platform Size Model Estimates

	Model 1		Model 2		Model 3	
Variable	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
Network Size Model						
Constant	4.3560*	0.0577	4.3570*	0.0613	4.0172*	0.0324
Marginal Network Value	0.2627*	0.1180	0.2580*	0.1143	0.5517*	0.1984
Retail Margin Model						
Constant	3.0186*	0.2279	2.7757*	0.2132	2.4422*	0.3821
Fruit Price	-1.0351*	0.1513	-0.9000*	0.1507	-0.5052*	0.2282
Veg Price	-0.1452*	0.0738	-0.0824	0.0680	-0.4967*	0.1206
Retail Wage	0.7122*	0.0751	0.6470*	0.0848	0.9139*	0.0856
Box Price	-0.8222*	0.0709	-0.7138*	0.0802	-0.8702*	0.0716
Utility Price	-0.2352	0.2121	-0.3042	0.3243	-0.8032*	0.1772
Fuel Price	-0.2357*	0.0253	-0.2003*	0.0385	-0.2260*	0.0211
Conduct Parameter	0.0926*	0.0470	0.0897*	0.0452	0.5519*	0.1769
R^2 / LLF / G	0.261		706.094		263.691	
R^2 Eq. 2	0.007					

Note: Model 1 is independent equations. Model 2 is NLSUR with no endogeneity controls. Model 3 is GMM with margin- and network-instruments. A single asterisk indicates significance at 5%.

Table 5. Counter-Factual Simulation of Indirect Network Effects

ϕ_N	Price	Std. Dev.	t-ratio	Network	Std. Dev.	t-ratio
100%	1.5968*	0.2855	2.4807	51.2097*	12.2843	12.6246
50%	1.5838	0.2790	1.2460	48.6327*	9.7849	6.8530
0	1.5710	0.2727		46.3802	7.7350	
-50%	1.5584	0.2666	-1.2566	44.4179*	6.1383	-7.5409
-100%	1.5459*	0.2607	-2.5234	42.7151*	5.0222	-15.0809

Note: Simulation conducted with estimates in table 4. A single asterisk indicates significant difference at a 5% level.

Table 6. Policy Simulations: Subsidizing Ugly Produce

η	Price	SD	t-ratio	Network	SD	t-ratio	Volume	SD	t-ratio
0%	1.5710	0.2727		46.3802	7.7350		17.4204	12.2361	
10%	1.6487	0.2731	5.3997	47.3121	0.1401	3.2322	21.2023	13.8297	5.4955
25%	1.7701	0.2509	14.4139	48.5055	0.4859	7.3581	27.9214	16.6419	13.6411
50%	1.9207	0.1886	28.2955	52.8411	8.0226	15.5564	39.7491	23.7332	22.4382
90%	2.1109	0.1566	46.0590	75.3317	3.5339	91.3504	53.6898	31.1292	29.0965

Note: t-ratio compares subsidy to 0% (base case). SD is the standard deviation.

A single asterisk indicates significance at a 5% level.

Figure 1. Users and Products by Week

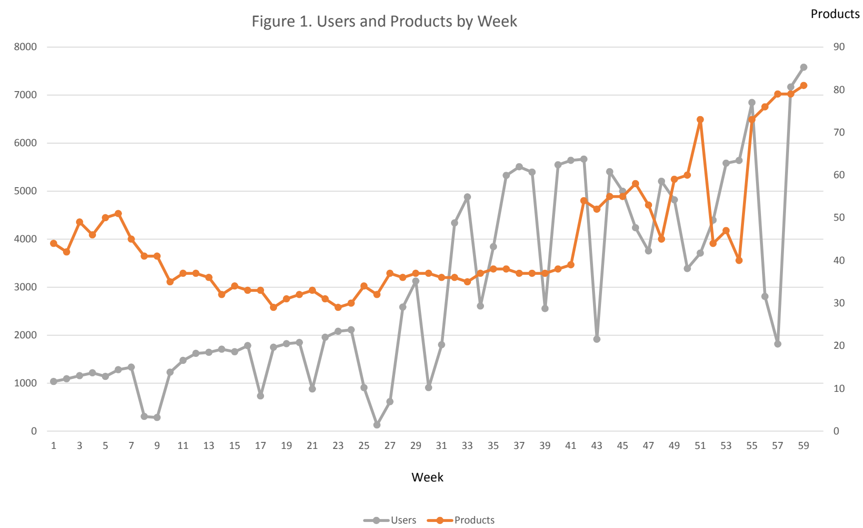

Figure 2. Gross Margin by Items