Testing Firm Conduct

Marco Duarte[*] Lorenzo Magnolfi[†] Christopher Sullivan[‡]

January 20, 2020

Abstract

Understanding the nature of firm competition is of first order importance. In most empirical environments, this require separately identifying marginal cost and conduct. While it is clear in an identification setup how to distinguish different models of conduct (Berry and Haile (2014)), the empirical literature has adopted several implementations whose properties have not been compared thus far. This paper seeks to understand the relative strengths and weaknesses of the choices available in the literature. We find that an approach based on Rivers and Vuong (2002), which takes into account the endogeneity of markups with respect to prices, is preferable in most environments. We perform Monte Carlo simulations to shed light on this, and examine test performance in two empirical applications. Our procedure delivers consistent and powerful inference, whereas some of the alternatives may be misleading.

1 Introduction

Understanding the nature of firm conduct is of central importance to academics and policy makers. In some settings, such as testing for collusion, how firms compete is a direct object of interest. In other settings, the nature of competition has ramifications for understanding the positive and normative effects of policy on market outcomes. The literature offers many theories of firm conduct, and may provide guidance on what type of conduct prevails. However, ultimately determining the exact nature of competition is an empirical question and requires using a practical method to distinguish among a menu of possible models.

The typical setting faced by researchers in IO is one in which we have data on prices and quantities in a market for differentiated products, but both firms' markups and marginal costs are unobserved. A static model of firm conduct generates equilibrium conditions which express prices as the sum of markups and marginal costs. Different models imply different values for markups. The goal of the researcher is thus to separately identify conduct and marginal cost. Bresnahan (1982), Lau (1982) and more recently Berry and Haile (2014) show that it is possible to test hypotheses

[*]Department of Economics, University of Wisconsin, Madison, WI 53706; duartefilho@wisc.edu.
[†]Department of Economics, University of Wisconsin, Madison, WI 53706; magnolfi@wisc.edu.
[‡]Department of Economics, University of Wisconsin, Madison, WI 53706; cjsullivan@wisc.edu.

on conduct in this setting so long as the researcher has instruments which shift and rotate the demand curve. Although the identification intuition is clear, there is no unanimity in the literature in how to implement it with a practical testing procedure. Currently, several different approaches to understanding the nature of conduct are used, yet to our knowledge, the strengths and weaknesses of these approaches have not been systematically investigated.

In this paper, we seek to provide guidance as to the best approach for evaluating firm conduct using data. We start by outlining our preferred procedure. This is a test in the spirit of Rivers and Vuong (2002). Importantly, this test accounts for the fact that markup terms - the key prediction of a conduct model - are jointly determined in equilibrium and are thus econometrically endogenous and easily accommodates the necessary excluded instruments.

We compare this procedure to other approaches popular in the literature. The first alternative approach we consider is a test based on the residual sum of squares for an OLS regression of implied marginal costs (prices minus markups) on cost shifters. Crucially, this test does not control for the underlying endogeneity of markups and thus is unreliable. We illustrate this feature in an example and provide a formal statement of what may fail. Monte Carlos illustrate that the problems we highlight are likely to be relevant in empirical settings.

There exist alternative testing procedures in the literature that also make use of excluded instruments. One such procedure is based on an auxiliary regression in which a researcher directly tests the exclusion restriction for the instruments under each model of conduct. The other procedure we consider estimates conduct and performs tests as a byproduct. While these procedures outline correctly specified tests under their maintained hypotheses, in Monte Carlos we find that our RV test compares favorably to them. In particular, our preferred test outperforms these alternatives in the presence of common features of empirical environments such as misspecification of demand or cost. We also find our procedure more interpretable than these approaches.

We show the performance of our proposed RV test in two empirical applications. Since it leverages on excluded instruments, the RV test helps avoid the bias arising from OLS regressions when using RSS - which can lead to misleading conclusions. Moreover, the RV test uses differences in values of the GMM objective function to assess which model constitutes a better approximation of the firm conduct that generates the data, whereas procedures based on auxiliary regressions may reject all hypotheses in the presence of even mild misspecification.

This paper discusses tools relevant to a broad literature seeking to understand firm conduct in the context of structural models of demand and supply. Investigating collusion is a prominent application (e.g. Porter (1983), Sullivan (1985), Bresnahan (1987), Slade (1992), Gasmi, Laffont and Vuong (1992), Genesove and Mullin (1998), Nevo (2001), Ciliberto and Williams (2014), Miller and Weinberg (2017), Bergquist and Dinerstein (2019), Sullivan (2020), Fan and Sullivan (2020), Eizenberg and Shilian (2019)). Other important applications include common ownership (Kennedy et al. (2017), Backus, Conlon and Sinkinson (2020)), vertical conduct (e.g. Villas-Boas (2007), Bonnet and Dubois (2010), Bonnet et al. (2013), Gayle (2013)), price discrimination (D'Haultfoeuille, Durrmeyer and Fevrier (2019)), price versus quantity setting (Feenstra and Levinsohn (1995), Varela and Viswanathan (2019), Magnolfi, Quint and Sullivan (2020)), post-merger internalization (Michel and Weiergraeber (2018)), and non-profit behavior (Duarte, Magnolfi and Roncoroni (2020)).

Many of the above examples involve contexts where the researcher does not observe marginal cost. However, there are some important exceptions: in Nevo (2001) for instance, accounting data on marginal costs are available. Other cases include Byrne (2015), Igami and Sugaya (2019), and Crawford et al. (2019). Such data are very helpful to assess firm conduct. While we discuss this briefly in Section 4, we are instead mainly concerned with the more common case where this data is not available to the researcher.

This paper is also related to econometric work on formally testing the relative fit of two alternative hypotheses. Vuong (1989) formulates a test of non-nested hypotheses for models estimated with MLE, and Rivers and Vuong (2002) extend the test to more general settings, including GMM. Hall and Pelletier (2011) further investigate the distribution of the Rivers and Vuong (2002) test statistic for the GMM case. More recently, work has been done to improve the asymptotic properties of the test. This includes Shi (2015) and Schennach and Wilhelm (2017). We make use of these recent advances in our applied environment.

The paper proceeds as follows. Section 2 describes the environment – a general model of firm conduct. Section 3 formulates our preferred testing procedure, based on Rivers and Vuong (2002). Section 4 describes common alternative in the literature, and compares our approach to them. Section 5 performs a set of Monte Carlos so to evaluate the relative performance across procedures. Section 6 discusses two empirical applications. Section 7 concludes.

2 Model of Firm Conduct

We consider the problem of distinguishing between two different types of firm conduct in a market for differentiated products. We assume that we observe a set of products $\mathcal{J}$ offered by firms across markets t. For each product j we observe price p_j, market share s_j, a vector of product characteristics x_j that affects demand, and a vector of characteristics w_j that affects the product's marginal cost. For each variable y we denote y_t as the vector of all values of that variable for products in market t. We assume that we either observe or can estimate the demand system $s_t = \sigma_t(p_t, x_t, \xi_t)$, [1] where ξ_t is the vector of unobserved product characteristics. Market equilibrium in market t is characterized by a system of first order conditions arising from firms' profit maximization problem:[2]

$$p_t = \Delta^0_t + c^0_t, \tag{1}$$

where Δ^0_t is the true vector of markup terms and c^0_t is the true vector of marginal costs. The true vectors of markups and marginal costs are unobserved by the econometrician.

[1] We return to discussing the implications for testing conduct of identifying demand and supply sequentially versus simultaneously later in the paper. [TBW]

[2] Although this is not the most general formulation, as it imposes a number of assumption on marginal revenues and marginal costs, we adopt it as it covers all the cases of practical interest in the literature.

The researcher can formulate alternative models of conduct m, which determine a corresponding vector of markups Δ_t^m. We can treat Δ_t^m as data since we have assumed that the researcher either observes or estimates all elements that determine markups such as demand elasticities.

Irrespective of the specific functional form of Δ_t^m, in oligopoly models this markup term is generally a function of both prices p_t and market shares s_t. In line with the literature, we assume that marginal costs are constant, and a linear function of observed cost shifters w_j and unobserved cost shifters ω_j.[3] Hence (1) for an assumed model m becomes:

$$p_{jt} = \Delta_{jt}^m(p_t, s_t) + w_{jt}\gamma + \omega_{jt}. \tag{2}$$

We follow the literature in assuming that w_{jt} is exogenous with respect to ω_{jt}. However Δ_{jt}^m is generally correlated with ω_{jt}. Note that this comes from two channels: one is direct, since Δ^m is a function of p, which is an explicit function of ω. The other is due to the fact that Δ^m is a function of shares s_t, which in turn depend on the unobservable ξ_t, which may be correlated with ω_t. In Appendix B we further explore the nature of endogeneity of markups within the context of a simple example.

Regardless of whether we want to estimate or test hypotheses based on equation (2), we need instruments for Δ_{jt}. Bresnahan (1982) and Berry and Haile (2014) provide intuition on what variation permits the researcher to separately identify conduct from cost, namely variation that shifts and rotates the demand curve. Common choices of instrumental variables of this type include the BLP instruments and the Gandhi and Houde (2019) instruments. Thus we assume that we have a vector of excluded instruments z_t that is orthogonal to ω_t.

Our aim is to use the model and the data to test two non-nested models of conduct, we can rewrite equation (2) as:

$$\begin{aligned} p_{jt} =& \Delta_{jt}^I\theta_1 + \Delta_{jt}^{II}\theta_2 + w_{jt}\gamma + \omega_{jt} && (3)\\ =& \Delta_{jt}\theta + w_{jt}\gamma + \omega_{jt} && (4) \end{aligned}$$

where $\theta = [\theta_1, \theta_2]'$ and the two models are

$$M^I : \theta_1 = 1, \theta_2 = 0 \qquad vs \qquad M^{II} : \theta_1 = 0, \theta_2 = 1. \tag{5}$$

This specification is general, allowing us to test many hypotheses on conduct found in the literature. We now discuss two canonical examples that fall into this framework.

Example 1: Test of Nature of Competition - In this case the researcher considers a pair of hypotheses on firm behavior that imply different functional forms of the markup function Δ_t^m. For example, we may be interested in testing hypotheses on the nature of firms' vertical relationships, or on whether firms compete in prices or quantities.

Example 2: Test of Internalization - We consider the case where the researcher is agnostic

[3] These assumptions are not necessary for our testing procedure to be valid, and could be relaxed.

on the degree to which a firm, when setting the price of a product, internalizes the effects of its own decision on the profits generated by other products. Standard examples of this would be collusion, intra-firm internalization, common ownership, nonprofit conduct.

Firms' first order conditions form the system of equations:

$$p_t = \left(\Omega(\tau) \odot \frac{\partial s_t}{\partial p_t}\right)^{-1} s_t + w_t\gamma + \omega_t, \tag{6}$$

where $\Omega(\tau)$ summarizes ownership and conduct assumptions. If the researcher has two distinct values of τ which represent economically interesting hypotheses, e.g. perfect collusion versus Nash, this case falls in our framework with

$$\Delta_t^m = \left(\Omega(\tau^m) \odot \frac{\partial s_t}{\partial p_t}\right)^{-1} s_t, \tag{7}$$

where different models constrain the ownership matrix to the values $\Omega(\tau^I)$ and $\Omega(\tau^{II})$.

3 Our Preferred Test

We propose a Rivers and Vuong (2002) testing procedure. In this framework, the null hypothesis of equivalent asymptotic fit of the two models is cast against two alternatives corresponding to the hypotheses of better asymptotic fit of model I and II respectively. For a finite sample of size n, the test is based on a lack-of-fit criterion Q_n^m for each model m. The test statistic is:

$$T_n^{I,II} = \frac{\sqrt{n}}{\hat{\sigma}_{I,II}}(Q_n^I - Q_n^{II}), \tag{8}$$

where $\hat{\sigma}^2_{I,II}$ is a consistent estimator of the variance of the difference in the measures of fit.

In what follows, we define lack of fit via the GMM objective function. Although the test can be defined for more general environments, in our setting we need to account for endogeneity, making GMM an appropriate choice. Specifically, given the excluded instruments z we defined above, we define sample moments

$$g_n^m(\gamma) = \frac{1}{n}[z\ w]'\omega^m(\gamma) \tag{9}$$

where $\omega_t^m(\gamma) = p_t - \Delta_t^m - w_t\gamma$. Hence,

$$Q_n^m = n g_n^m(\hat{\gamma}^m)' W_n g_n^m(\hat{\gamma}^m) \tag{10}$$

where W_n is a positive semi-definite weighting matrix, and $\hat{\gamma}^m$ is the GMM estimate of γ under model m.

Following Rivers and Vuong (2002) - RV in what follows - the null hypothesis for the test is that the two models M^I and M^{II} are $\sqrt{n}-$asymptotically equivalent, or

$$H_0: \lim_{n\to\infty} \sqrt{n}\{Q_n^I - Q_n^{II}\} = 0 \tag{11}$$

Relative to this null, we can define two alternative hypotheses corresponding to cases of $\sqrt{n}-$asymptotically better fit of one of the two models:

$$H_I: \lim_{n\to\infty} \sqrt{n}\{Q_n^I - Q_n^{II}\} = -\infty \qquad \text{and} \qquad H_{II}: \lim_{n\to\infty} \sqrt{n}\{Q_n^I - Q_n^{II}\} = +\infty \tag{12}$$

Rivers and Vuong (2002) show that the distribution of the test statistic $T_n^{I,II}$ is standard Normal under the null.

There are a few practical considerations in implementing our preferred test, for which we provide guidance here. To compute the test statistic, values of Q_n^m for $m = I, II$ and an estimator for the variance $\hat{\sigma}^2_{I,II}$ are necessary. These objects in turn depend on the choice of the weight matrix W_n. Hall and Pelletier (2011) derive formulas for the estimator $\hat{\sigma}^2_{I,II}$ under two commonly used weight matrices - identity and $([z\ w]'[z\ w])^{-1}$. We will use the latter weight matrix in what follows and reproduce the formula for $\hat{\sigma}^2_{I,II}$ in an appendix TBW.

[Also TBW: Discuss more recent econometric issues with RV: Schennach and Wilhelm, Shi.]

4 A Comparison with Other Approaches in the Literature

In this section we compare our preferred RV approach as formulated above with commonly used procedures used in the literature. This includes a goodness of fit test based on residual sum of squares (RSS), a procedure based on an auxiliary regression similar to a Anderson and Rubin (1949) test, and an estimation-based type of procedure where parameters governing conduct are directly estimated. Supported by Monte Carlo simulations, we'll show how these different procedure compare with our approach. We maintain our preference for the RV approach for several reasons: this procedure directly compares the two hypotheses on conduct to each other, allows for misspecification of the model, facilitates the inclusion of instruments, and is easily interpretable. The above approaches are particularly useful in the context we discuss in Section 2 - where the marginal costs are unobserved by the econometrician. An important caveat is that in some environments the researcher may have access to external data on marginal costs or markups, which then can be used to directly discriminate between models of conduct. We discuss this fourth approach at the end of the section.

4.1 RSS Goodness of Fit

The test that's closest in spirit to our preferred procedure is a goodness of fit test based on RSS. To perform this test, notice that once we impose model m, we have the following linear model in w:

$$p_t - \Delta_t \theta^m = w_t \gamma + \omega_t. \tag{13}$$

Under the imposed model, $p_t - \Delta_t \theta^m$ is observed, and we can obtain an OLS residual $\hat{\omega}_t^m$ from the regression (13). One then constructs for model m a measure of fit $Q_n^{RSS,m} = \frac{1}{n}\sum_t (\hat{\omega}_t^m)'(\hat{\omega}_t^m)$, and then computes a test statistic analogous to (8):

$$\tilde{T}_n^{I,II} = \frac{\sqrt{n}}{\hat{\sigma}_{I,II}^{RSS}}(Q_n^{RSS,I} - Q_n^{RSS,II}), \tag{14}$$

where $\hat{\sigma}_{I,II}^{RSS}$ is a consistent estimator of the standard deviation of the difference in the RSS measures of fit. The formulation of the hypotheses is identical to what we have described in Section 3 above. This procedure has several virtues: it's simple to implement, as it involves little more than estimating two OLS regressions and directly compares the two hypotheses of interest. It has also been popular in the applied literature in Industrial Organization: for example, Villas-Boas (2007), Bonnet and Dubois (2010) and Bonnet et al. (2013) use it to test hypotheses on conduct in vertical markets.

However, this approach may be misleading. To see this point, consider (3)

$$p_{jt} = \Delta_{jt}\theta + w_{jt}\gamma + \omega_{jt}, \tag{15}$$

and suppose that the data are generated according to some true parameter values (θ_0, γ_0). Suppose that we estimate equation (15) with OLS, obtaining the estimates $\hat{\theta}_{OLS}$ and $\hat{\gamma}_{OLS}$. As we already noticed, because markups Δ are endogenous, this does not generate consistent estimates of the true parameters. Hence, we can show that there always exists a value of θ that offers a better RSS fit than the true value $\theta = \theta_0$. To do so, consider the two models:

$$M_I : \theta = \theta_0 \qquad vs \qquad M_{II} : \theta = \hat{\theta}_{OLS} \tag{16}$$

and for either model m we can define the constrained OLS estimate of γ as

$$\hat{\gamma}^m = \hat{\gamma}_{OLS} - \frac{V_{2,1}}{V_{1,1}}(\hat{\theta}_{OLS} - \theta^m) \tag{17}$$

where V_{ij} denotes elements (i, j) of the variance-covariance matrix of the OLS estimators $(\hat{\theta}_{OLS}, \hat{\gamma}_{OLS})$. Since we have defined $\theta^{II} = \hat{\theta}_{OLS}$, then $\hat{\gamma}^{II} = \hat{\gamma}^{OLS}$. Let $\hat{\omega}_{jt}(\theta, \gamma) = p_{jt} - \Delta_{jt}\theta - w_{jt}\gamma$ and notice that

$$nQ_n^{RSS,II} = \sum \hat{\omega}_{jt}(\theta^{II}, \hat{\gamma}^{II})^2 \tag{18}$$

$$= \sum \hat{\omega}_{jt}(\hat{\theta}_{OLS}, \hat{\gamma}_{OLS})^2 \tag{19}$$

$$< \sum \hat{\omega}_{jt}(\theta, \gamma)^2, \qquad \forall\, (\theta, \gamma) \neq (\hat{\theta}_{OLS}, \hat{\gamma}_{OLS}) \tag{20}$$

from which it follows that $Q_n^{RSS,II} < Q_n^{RSS,I}$: the incorrect value of $\theta = \hat{\theta}_{OLS}$ has thus better RSS fit than the truth $\theta = \theta_0$. [4]

[4] Even in the case in which Δ and w are uncorrelated - thus making $\hat{\gamma}^m = \hat{\gamma}_{OLS}$ for any value of θ - the point above still stands and the true value $\theta = \theta_0$ does not minimize the RSS loss function.

The feature of the test we just highlighted may seem benign: after all, sampling noise may always generate better fit for misspecified models. What makes this property problematic, however, is that as the sample size goes to infinity, misspecified models still fit better than the true one according to this RSS-based test, since $plim\hat{\theta}_{OLS} = \theta_{OLS} \neq \theta_0$. Hence, even in large samples, the RSS-based testing procedure we just outlined may generate misleading inference.

In fact we can prove that for any two values of θ, the one with better asymptotic RSS fit is the one that is closest to the plim of the OLS estimate of θ.

Proposition 1. *If* $||\theta^I - \theta_{OLS}|| < ||\theta^{II} - \theta_{OLS}||$, *then* $plim(Q_n^{RSS,I} - Q_n^{RSS,II}) < 0$.

The proof of Proposition 1 is provided in Appendix A. Whether the RSS-based procedure leads to rejecting the null in favor of the wrong hypothesis depends on the value of the variance term at the denominator of the test statistic. However, Proposition 1 shows that the denominator of the test statistic may have the wrong sign, and the Monte Carlo simulations in the next section shows that – in a reasonable empirical environment – this may actually lead to reject in favor of misspecified models.

While testing based on RSS may appear to only rely on the OLS fit of (13), where all variables on the right hand side are exogenous negating the need for instruments, Proposition 1 shows that the results of RSS-based tests are driven by the relative distance of the two models from the OLS estimates of (15). Due to the fact that Δ is generally correlated with ω, the OLS estimates of (15) are inconsistent, thus potentially skewing the test.

An obvious solution is to include instruments as in the RV procedure of Section 3. Suppose we were to estimate (15) using valid instruments for markups. Then the plim of the resulting estimator of θ would be the true value θ_0. The GMM objective functions Q_n^m measure model fit in a way that corresponds to this estimation exercise, and do not incur in the issues we have discussed for RSS, as the following proposition holds:

Proposition 2. *If* $||\theta^I - \theta_0|| < ||\theta^{II} - \theta_0||$, *then* $plim(Q_n^I - Q_n^{II}) < 0$.

The proof of Proposition 2 is provided in Appendix A.

This proposition motivates the use of our RV testing procedure outlined in Section 3. While maintaining the advantages of RSS testing, the use of instruments avoids the possibility of misleading inference.

A final point: in the literature, this RSS based approach has been called an RV test. It is in fact not an RV test, since it compares two models with different dependent variables. This can be easily seen from (13): different models m imply different markups and hence different marginal costs.

4.2 Auxiliary Regression Approach

In this approach the researcher defines, for each of the two models m an equation:

$$p_{jt} - \Delta_{jt}^m = w_{jt}\gamma + z_{jt}\pi + \eta_{jt}, \tag{21}$$

and then performs the test of the null hypothesis that $\pi = 0$ with either an F-test (Likelihood Ratio approach) or a Wald test (t-test if z is a scalar). Since $\pi = 0$ is a necessary condition for model m to be true, a rejection of $\pi = 0$ results in a rejection of model m. If implemented as an F-test, this approach is similar to an Anderson and Rubin (1949) testing procedure, which allows to test hypotheses in linear models with endogeneity in the presence of possibly weak instruments. A desirable feature of this test is that it can accommodate misspecification in the marginal cost function, as long as the exclusion restrictions are valid. Recent examples of this test in the applied literature are Backus, Conlon and Sinkinson (2020) and Bergquist and Dinerstein (2019).

There is a fundamental difference in how the models are compared in this approach relative to our RV test. The RV test concludes that either both models fit equally well asymptotically, or that one fits better relative to the other. This allows for both models to be misspecified - which is likely to be the case in empirical work. The auxiliary regression approach instead - separately for each model - tests for a necessary condition for the model to be true against an alternative that rejects the model. Hence, it may happen that the null is not rejected for both models, which is not immediately interpretable. Moreover, even a modest amount of misspecification of demand (and hence of markups) could lead to the rejection of a model that fits well. We will further explore these points in our Monte Carlo simulations in Section 5.

4.3 Estimation Based Approach

Testing can also be conducted as a by-product of estimation of the parameters of the model. Estimation can take two forms based on the two examples in Section 2. In example 1, the estimating equation is analogous to (3) and the researcher can directly test hypotheses on the θ parameter corresponding to two models. In example 2, the estimating equation is instead:

$$p_{jt} = \Delta_{jt}(\tau) + w_{jt}\gamma + \eta_{jt}, \tag{22}$$

and the researcher tests hypotheses on the τ parameter. If estimation is conducted in a GMM framework, the hypotheses in (39) and (40) can be tested with a J-test (Likelihood Ratio approach) or with a Wald test. The test can be performed by checking whether each model m implies parameters that lie within the confidence interval of the estimates. Recent examples in the applied literature of estimation that lend themselves to this testing procedure include Ciliberto and Williams (2014), Miller and Weinberg (2017) and Michel and Weiergraeber (2018).If instruments are strong, this test has more power than the ones obtained under the auxiliary regression approach.

4.4 External Marginal Cost or Markup Data

The researcher may either observe an exact or a noisy measure of marginal costs or markups. For instance, accounting data may be available as in Nevo (2001). In this case, the researcher can proceed by comparing directly the markups Δ^m or marginal costs c^m implied by each model m with the data, thus obviating the need for the testing procedures described above. Even in the absence of observed marginal costs or markups, there could be something useful for the researcher to learn before the test by examining the distributions of markups or costs implied by the two candidate models. For instance, a model may have implications that are inconsistent with logic or institutional details, such as many observations having negative implied marginal costs. Such model is clearly implausible. While not a formal test, this type of inspection could be useful in formulating and assessing the validity of models. Conversely, both of the models under consideration may imply very similar distributions of markups and marginal costs. A caveat for all the testing procedures described above is that in this case testing will be extremely difficult and could be misleading, as we will see in the Monte Carlo simulations. For all these reasons, we believe it is best practice for the researcher to visually inspect the distribution of markups before proceeding with the formal test.

5 Monte Carlo Evidence

We present in this section Monte Carlo evidence meant to illustrate the performance of the testing procedures discussed in Sections 3 and 4. We start by describing the setup for the simulation.

5.1 Setup

Market Environment - We generate data from a static model of demand and supply in the class described in Section 2. The demand side follows the standard mixed logit formulation. In market t, consumer i chooses among $\mathcal{J}_t = \{1, ..., J_t\}$ products and has an indirect utility from product j of

$$U_{ijt} = x_{jt}\beta + p_{jt}\alpha + \xi_{jt} + \mu_{ijt} + \varepsilon_{ijt} \tag{23}$$

where utility contains two consumer-product specific components: μ_{ijt}, which depends on product attributes (x, p) and consumer attributes $\upsilon \sim N(0, 1)$ according to

$$\mu_{ijt} = x_{jt}\upsilon_{itx}\sigma_x + p_{jt}\upsilon_{itp}\sigma_p.$$

and a shock $\varepsilon_{ijt} \sim T1EV$. The parameters governing consumer preferences are α, β and σ. For a given value of these parameters, the model implies a matrix $D_p \equiv [\partial s_j/\partial p_k]$ of partial derivatives of product j's market share with respect to prices.

On the supply side we make use of three models of firm conduct that determine the form of markups Δ^m in equation (2): Nash-Bertrand price setting (B), Nash-Cournot quantity setting (C), and joint profit maximization price setting (J). As in equation (2), costs are assumed to be the sum of a linear index of the observed cost shifters w and an unobserved component ω. For the three models we consider, markups are respectively:

$$\Delta^B_{jt} = (\Omega \odot D_p)^{-1} s_{jt} \tag{24}$$

$$\Delta^C_{jt} = (\Omega \odot D_p^{-1}) s_{jt} \tag{25}$$

$$\Delta^J_{jt} = (\Omega^J \odot D_p)^{-1} s_{jt}, \tag{26}$$

where Ω is the ownership matrix where element (i, j) is an indicator for products i and j being produced by the same firm, and Ω^J is a matrix of ones.

Parametrization - We choose a parametrization loosely based on Conlon and Gortmaker (2019) as baseline. The parameter values create realistic outside shares in Bertrand, around (0.8,0.9). In the baseline, we have $N_m = 200$ markets and $N_f = 10$ firms. We allocate products and firms randomly to markets. Each market can have either 3, 4 or 5 firms and those firm can have either 2, 5 or 8 products. Firm-product pairs are drawn first and then allocated to markets. The vector of product characteristics x contains a constant and a scalar $\tilde{x}$. The vector of cost shifters w includes a constant, $\tilde{x}$ and a scalar $\tilde{w}$. Both $\tilde{x}$ and $\tilde{w}$ are independently drawn from a standard uniform distribution. In our baseline specification we set the following parameter values:

$$\begin{aligned}
\text{Preference parameters:} \quad & \sigma_x = 3, \quad \sigma_p = 0.25, \quad \alpha = -1, \quad \beta = (-6, 6) \\
\text{Cost parameters:} \quad & \gamma = (2, 1, 0.2) \\
\text{Unobserved shocks:} \quad & (\xi_{jt}, \omega_{jt}) \sim N(0, \Sigma), \quad \sigma^2_\xi = \sigma^2_\omega = 0.2 \text{ and } \sigma_{\xi\omega} = 0.
\end{aligned}$$

Given parameter values, we draw observed x, w and unobserved shocks for each of the $NS = 100$ simulations and compute endogenous prices and quantities as the solution of the market share equations and first order conditions for each model m.

Instruments - We construct two types of instruments for the Monte Carlo exercise. One version of BLP instruments and one of differentiation instruments. For product-market j, let O_j be the set of products other than j sold by the firm that produces j and let C_j be the set of products produced by rival firms. The instruments are:

$$z^{BLP}_j = \left[\sum_{i \in O_j} 1[i \in O_j] \quad \sum_{i \in C_j} 1[i \in C_j] \quad \sum_{i \in O_j} x_i \quad \sum_{i \in C_j} x_i \right]$$

$$z^{Diff}_j = \left[\sum_{i \in O_j} d^2_{ji} \quad \sum_{i \in C_j} d^2_{ji} \quad \sum_{i \in O_j} 1[|d_{ji}| < sd(d)] \quad \sum_{i \in C_j} 1[|d_{ji}| < sd(d)] \right]$$

where $d_{ij} \equiv \tilde{x}_i - \tilde{x}_j$ and $sd(d)$ is the standard deviation across the distance vector.

5.2 Comparison between RV Test and RSS Test

In this subsection we show how endogeneity leads RSS-based tests to generate misleading conclusions. We do so comparing a pair of data generating processes corresponding to models of firm behavior: Nash-Bertrand (B) versus joint profit maximization (J). For both of these DGPs, we compare the RV test with the RSS test. In what follows, we use the differentiation IVs (z^{Diff}) to specify the RV test.

We start by simulating data for the case where the true model of conduct is Nash Bertrand competition. For each of the 100 simulations of the data, we perform both our preferred RV test and the RSS test of fit. Row (1) of Table 1 summarizes the results for our baseline specification. For the RSS test, we see that the null hypothesis of equal fit cannot be rejected in 1 percent of the simulations. Meanwhile, the RSS test rejects in favor of the hypothesis that the Bertrand model offers better fit 99 percent of the time. The RV test performs similarly well, rejecting the null in favor of better fit of the Nash Bertrand model in every simulation.

Table 1: Test of Bertrand vs Joint Profit Maximization
DGP = Bertrand

	RSS Test			RV Test		
Version	H_I: B	H_0	H_{II}: J	H_I: B	H_0	H_{II}: J
(1)	0.99	0.01	0.00	1.00	0.00	0.00
(2)	0.08	0.37	0.55	0.92	0.08	0.00
(3)	0.00	0.12	0.88	0.94	0.06	0.00
(4)	0.00	0.00	1.00	0.76	0.24	0.00

(1) baseline: $\sigma^2_\omega = 0.2,\ \sigma_{\xi\omega} = 0,\ \sigma_x = 3$

(2) $\sigma^2_\omega = 1,\ \sigma_{\xi\omega} = 0,\ \sigma_x = 3$

(3) $\sigma^2_\omega = 1,\ \sigma_{\xi\omega} = 0.4,\ \sigma_x = 3$

(4) $\sigma^2_\omega = 1,\ \sigma_{\xi\omega} = 0.4,\ \sigma_x = 2$

Table reports, for each testing environment and for each hypothesis, the fraction of simulations in which the test either did not reject (for H_0) concluded in favor of that hypothesis (for either H_m). Rows (2) - (4) augment the baseline in the described way. Unless specified, parameters remain the same as in baseline.

We next augment the baseline to highlight important features of the RV test relative to the RSS test. First, we increase the amount of endogeneity in our specifications by increasing the variance of ω. In Row (2), we see the main problem with the RSS test highlighted above. While the RV test concludes in favor of the Bertrand model in 92 percent of simulations, the RSS test concludes in favor of the Bertrand model in only 8 percent of simulations. Meanwhile the RSS test rejects the null in favor of the alternative of superior fit of joint profit maximization, the wrong model, in 55 percent of simulations. Row (3) adds additional endogeneity through the indirect channel, by increasing $\sigma_{\omega\xi}$. While the performance of the RV test is largely unchanged, RSS rejects the null in favor of the wrong model in 88 percent of the simulations while never rejecting the null in favor of the correctly specified model. In this simple setting, the RSS test is highly susceptible to the degree of endogeneity and one cannot perform testing without valid instruments for the markups.

In addition to endogeneity, the tests may struggle to reject in favor of the correct hypothesis if the distribution of markups under the two models are sufficiently close. In row (4) we seek to compress to distribution of markups by reducing the amount of consumer heterogeneity in preferences and therefore making the substitution patterns increasingly similar to those which would obtain from a standard logit model.[5] In row (4), we maintain the level of endogeneity in (3) while reducing the value of σ_x from 3 to 2. Here we see that while the RV test rejects in favor of Bertrand competition in 76 percent of simulations, the RSS test rejects in favor of joint profit maximization in every simulation.

For the next set of simulations, we generate the data from the model of joint profit maximization. We then test the fit of the model of joint profit maximization against that of Nash Bertrand using both the RV test and the RSS test. Table 2 presents the results of these simulations for two of the environments considered in Table. We first see that in the baseline specification the RV test performs similarly well in this case to the previous case where the DGP was Bertrand. However, we see that for both the baseline and an environment with endogeneitythe RSS test always rejects the null in favor of the correct model of joint profit maximization.[6] Compare the the result of Row (3),

Table 2: Test of Bertrand vs Joint Profit Maximization
DGP = Joint Profit Maximization

	RSS Test			RV Test		
Version	H_I: B	H_0	H_{II}: J	H_I: B	H_0	H_{II}: J
(1)	0.00	0.00	1.00	0.00	0.03	0.97
(3)	0.00	0.00	1.00	0.00	0.17	0.83

(1) baseline: $\sigma^2_\omega = 0.2,\ \sigma_{\xi\omega} = 0,\ \ \sigma_x = 3$

(3) $\sigma^2_\omega = 1,\ \sigma_{\xi\omega} = 0.4,\ \ \sigma_x = 3$

Table reports, for each testing environment and for each hypothesis, the fraction of simulations in which the test either did not reject (for H_0) concluded in favor of that hypothesis (for either H_m). Row (3) augments the baseline in the described way. Unless specified, parameters remain the same as in baseline.

where we introduce endogeneneity, in both tables. In Table 1, the endogeneity resulted in the RSS test rejecting in favor of joint profit maximization (the wrong hypothesis) 88 percent of the time. In Table 2, the endogeneity results in the RSS test concluding in favor of joint profit maximization 100 percent of the time. It becomes apparent that in this setup the nature of the bias resulting from endogeneity pushes the RSS test statistic, regardless of the DGP, towards rejecting in favor of joint profit maximization. Thus, while RSS superficially seems to perform well in Table 2, it does so for a reason that undermines the credibility of the test.

[5] In the logit model, the markups are the same for all product sold by a given firm in a market. Thus, a logit model relative to a mixed logit model will have less within market variation in the markups.

[6] The same happens for the testing environments corresponding to rows (2) and (4) in Table 1.

5.3 Comparison among RV Test, Auxillary Regression Approach and Estimation Based Approach

In this subsection, we compare the performance of the RV test with the other commonly used alternatives: an auxiliary regression approach (AR) and an estimation based approach (EB).[7] We do so by comparing a pair of data generating processes corresponding to models of firm behavior: Nash-Bertrand (B) versus joint profit maximization (J). In what follows, we use the differentiation IVs (z^{Diff}) to specify the RV test.

All three approaches are designed to account for endogeneity and thus perform well not only in our baseline specification, but also in one where we increase both channels of endogeneity. In Table 3, we report the results from comparing the performance of the three testing procedures in the same environment considered in row 3 of Tables 1 and 2.

We report in the table the fraction of simulations for which a specific test concludes in favor of an outcome. Notice that this statement has different meanings in the testing environments. For instance, when RV concludes in favor of B only, this is rejecting the null of equal fit in favor of an alternative of better fit of B. Meanwhile for AR or EB, concluding for B only means not rejecting the null that B is the true model while rejecting the null that J is the true model pitted against all possible alternatives. As expected, all procedures perform well in this setting.

TABLE 3: Test Comparison - $\sigma^2_\omega = 1$, $\sigma_{\omega\xi} = 0.4$

	DGP B			DGP J		
Models	RV	AR	EB	RV	AR	EB
B only	0.94	0.93	0.96	0.00	0.00	0.01
J only	0.00	0.00	0.00	0.83	0.88	0.92
$B = J$	0.06			0.17		
Both B and J		0.01	0.01		0.06	0.02
Neither B nor J		0.06	0.03		0.06	0.05

This table compares the performance of the three testing procedures: our RV procedure, the auxiliary regression based approach (AR) and the estimation based approach (EB) for the two models: Bertrand (DGP B) and joint profit maximization (DGP J). Rows correspond all possible outcomes across these tests. Columns correspond to a particular testing environment applied to data generated by a specific DGP. We report in the table, the fraction of simulations for which a specific test concludes in favor of a specific outcome.

The setting above was an ideal testing environment in that (1) there was no misspecification of either demand or marginal cost, and (2) demand parameters were not estimated but fixed to their true values. In empirical work instead, these are likely to be prominent issues. Thus, it is interesting to compare the relative performance of these three approaches in settings where the model is misspecified, or demand is estimated.

Misspecification of Cost - Here we explore how the three tests perform in a setting where demand is known but the marginal cost function is misspecified. In our baseline model, marginal

[7] Details for how we implement these alternative approaches can be found in Appendix - TBW.

cost is a function of two observed cost shifters x (which also enters demand), and w (which is excluded from demand). We misspecify marginal cost in two ways, by separately leaving out one of the variables from the supply equation.

Table 4 reports the results for simulations where w is omitted from the supply specification. First, the RV and EB test vastly outperform the AR procedure. Although the AR procedure is a direct test of the exclusion restriction and should not, in large sample, suffer from a misspecification of marginal cost which does not invalidate the exclusion restriction, we see here that in finite sample, misspecification of this form leads the AR procedure to reject both null hypotheses in 28 percent of simulations.

Table 4: Cost misspecification - w omitted

	DGP B			DGP J		
Models	RV	AR	EB	RV	AR	EB
B only	1.00	0.72	1.00	0.00	0.00	0.00
J only	0.00	0.00	0.00	0.96	0.72	0.97
$B = J$	0.00			0.04		
Both B and J		0.00	0.00		0.00	0.03
Neither B nor J		0.28	0.00		0.28	0.00

This table compares the performance of the three testing procedures: our RV procedure, the auxiliary regression based approach (AR) and the estimation based approach (EB) for the two models: Bertrand (DGP B) and joint profit maximization (DGP J). Rows correspond all possible outcomes across these tests. Columns correspond to a particular testing environment applied to data generated by a specific DGP. We report in the table, the fraction of simulations for which a specific test concludes in favor of a specific outcome.

Table 5 reports the results for simulations where x is omitted from the supply specification. This is a more extreme form of misspecification as it invalidates the moment condition along which all three tests rely. However, it is empirically relevant in that instrument validity, unlike weakness, cannot be checked by the researcher. In this challenging environment, the RV approach manages to perform well, concluding in favor of the correct model in 71 percent and 100 percent of the simulations respectively. The AR procedure, relying on a different formulation of the null hypothesis, rejects both models in 97 percent of simulations, a result that while accurate is uninformative to the researcher.

Table 5: Cost Misspecification - x omitted

	DGP B			DGP J		
Models	RV	AR	EB	RV	AR	EB
B only	0.71	0.03	0.47	0.00	0.00	0.00
J only	0.14	0.00	0.05	1.00	0.03	0.42
$B = J$	0.15			0.00		
Both B and J		0.00	0.00		0.00	0.00
Neither B nor J		0.97	0.48		0.97	0.58

This table compares the performance of the three testing procedures: our RV procedure, the auxiliary regression based approach (AR) and the estimation based approach (EB) for the two models: Bertrand (DGP B) and joint profit maximization (DGP J). Rows correspond all possible outcomes across these tests. Columns correspond to a particular testing environment applied to data generated by a specific DGP. We report in the table, the fraction of simulations for which a specific test concludes in favor of a specific outcome.

Misspecification of Demand - Here we explore how the three tests perform in a setting where demand is misspecified. We misspecify demand cost in two ways, by separately misspecifying the variance of either the random coefficient on x or p.

Table 6 reports the results for simulations where σ_x, the variance of the random coefficient on product characteristics, is set to a value of 2 as opposed to its true value of 3. In this case, the RV procedure performs very well, rejecting the null in favor of the correct model in 99 percent and 95 percent of cases for the two DGPs. In contrast, the auxiliary regression based approach struggles, rejecting both models in a vast majority of simulations (97 percent and 85 percent for the two DGPs respectively). The estimation based approach does not reject the right model in 52 and 81 percent of the simulations respectively, but offers a performance that significantly lags behind our RV procedure.

Table 6: Demand misspecification - $\hat{\sigma}_x = 2$

	DGP B			DGP J		
Models	RV	AR	EB	RV	AR	EB
B only	0.99	0.03	0.52	0.00	0.00	0.00
J only	0.00	0.00	0.00	0.95	0.15	0.81
$B = J$	0.01			0.05		
Both B and J		0.00	0.02		0.00	0.14
Neither B nor J		0.97	0.46		0.85	0.05

This table compares the performance of the three testing procedures: our RV procedure, the auxiliary regression based approach (AR) and the estimation based approach (EB) for the two models: Bertrand (DGP B) and joint profit maximization (DGP J). Rows correspond all possible outcomes across these tests. Columns correspond to a particular testing environment applied to data generated by a specific DGP. We report in the table, the fraction of simulations for which a specific test concludes in favor of a specific outcome.

Table 7 reports the results for simulations where σ_p, the variance of the random coefficient on price, is set to a value of 0 as opposed to its true value of 0.25. As above, RV outperforms the other two procedures, rejecting the null in favor of the correct model in 100 percent and 95 percent of cases for the two DGPs.

Table 7: Demand Misspecification - $\hat{\sigma}_p = 0$

	DGP B			DGP J		
Models	RV	AR	EB	RV	AR	EB
B only	1.00	0.40	0.81	0.00	0.00	0.00
J only	0.00	0.00	0.00	0.95	0.87	0.94
$B = J$	0.00			0.05		
Both B and J		0.00	0.00		0.02	0.06
Neither B nor J		0.60	0.19		0.11	0.00

This table compares the performance of the three testing procedures: our RV procedure, the auxiliary regression based approach (AR) and the estimation based approach (EB) for the two models: Bertrand (DGP B) and joint profit maximization (DGP J). Rows correspond all possible outcomes across these tests. Columns correspond to a particular testing environment applied to data generated by a specific DGP. We report in the table, the fraction of simulations for which a specific test concludes in favor of a specific outcome.

Estimation of Demand - [To Be Added, initial results suggest that estimating demand in finite sample dramatically worsens the performance of AR and EB relative to RV. In this section, we also seek to explore the relative performance of these tests when demand is estimated before testing supply versus the case in which demand and supply are jointly considered. The literature has made strong and contradictory claims regarding sequential vs simultaneous estimation in models where conduct is unknown and to our knowledge, this has not been systematically explored. We aim to do that here.]

6 Empirical Illustration

6.1 Testing for collusion

[To be written, based on Sullivan (2020).]

6.2 Cooperative Conduct

Duarte, Magnolfi and Roncoroni (2020) investigate the competitive conduct of a large Italian grocery retailer, Coop Italia. This firm is a consumer cooperative, and thus should internalize the welfare of its consumer-members when setting prices in its stores. On the other hand, weak governance may allow managers to pursue other goals. For instance, managers may want to reinvest profits to grow the firm and increase their perks consumption, thus resulting in profit-maximizing conduct, since Hence, it is interesting to empirically test for two theories on Coop's objectives: perfect internalization of consumer welfare, whereby the firm is entirely run in the interest of its consumer-members (subject to the constraint of not making losses), and profit maximization.

The supply-side environment in Duarte, Magnolfi and Roncoroni (2020) fits in the framework examined in equation 3, with the two models of pure profit maximization and internalization consumer welfare implying markup terms equal to Δ^B_{jt}, the Bertrand-Nash markup, and zero respectively. We can thus characterize the pricing decision of firm (supermarket) j in market t as:

$$p_{jt} = \Delta^I_{jt}\theta_1 + \Delta^{II}_{jt}\theta_2 + w_{jt}\gamma + \omega_{jt} \tag{27}$$

$$= \Delta^B_{jt}\theta_1 + w_{jt}\gamma + \omega_{jt} \tag{28}$$

Hence, model M^I implies a value $\theta_1 = 1$, whereas model M^{II} implies a value $\theta_1 = 0$. Although Duarte, Magnolfi and Roncoroni (2020) estimate this equation and perform testing using an estimation based approach, the other testing approaches discussed in this paper are available for their application. We report in Table 8 the test statistics and critical values for the four testing procedures that we examine in this paper.

In this application the test based on RSS and our RV procedure both lead to the same conclusion, since the negative values of the test statistic (of -12.7 and -4.7, respectively) reject the null of equal fit of the two models in favor of superior fit of M_I, pure profit maximization. In light of our Propositions 1 and 2 and of the estimation results for equation 3, this is not surprising: the OLS point estimate of the parameter θ_1 is 0.82, much closer to the value of 1 implied by M_I than to the value of zero implied by M_{II}. Similarly, the IV estimate - obtained using the same instruments that we employ to construct the RV test - is 1.02, almost identical to the value implied by M_I.

The auxiliary regression based method, however, rejects both models – although model M_{II} is rejected with a much higher test statistic. This result is best interpreted in light of the previous section, where we have shown in Monte Carlo simulations that even a small amount of misspecification in demand or marginal cost can lead the AR method to reject both models. If we believe that a model can at best represent a credible approximation of reality, as opposed to being literally true, this feature of the AR test is unappealing. Finally, the estimation based approach rejects M_{II} and does not reject M_I, thus falling in line with the results of the RV test.

TABLE 8: Coop Conduct Test - M_I: $\theta_1 = 1$ M_{II}: $\theta_1 = 0$

	RSS	RV	AR		EB	
			M_I	M_{II}	M_I	M_{II}
Statistic	-12.70	-4.70	4.73	23.38	0.05	3.85
critical value	±1.96	±1.96	2.10	2.10	±1.96	±1.96

Table reports values of the test statistic for the four testing procedures we describe in the paper: RSS-based, our preferred RV, auxiliary regression (AR), and estimation based (EB). For each procedure we report one critical value (in the cases of RSS and RV) or two critical values corresponding to tests where the nulls are models M_I and M_{II}, respectively.

7 Conclusion

In this paper, we explore the relative performance of different testing procedures in an empirical environment encountered often by IO economists: comparing two competing hypotheses of firm conduct. We find that a procedure based on Rivers and Vuong (2002) offers a highly desirable set of features. First, it accommodates endogeneity. This is in sharp contrast to a similar procedure used in the literature bases on RSS fit. We prove that the RSS test can be misleading in a way that the RV test cannot if the researcher has valid instruments. In a set of Monte Carlo simulations, we show that this concern is empirically relevant as it can lead RSS to conclude in favor of the wrong model.

Second, the RV test accommodates misspecification of both demand and marginal cost. This is in contrast to alternative testing procedures which accommodate endogeneity. We show in simulations that when cost is misspecified, the RV test vastly outperforms a testing procedure based on an auxiliary regression. When demand is misspecified, the gap in performance between the RV test and alternative procedures is even larger.

Finally, in two empirical applications, we implement the RV test. We show that the procedure is practical and highlight the sensitivity of endogeneity. In the first example, whereas the RV test concludes in favor of full collusion an RSS test would strongly reject this model in favor of Nash Bertrand behavior. In the second example, using an AR test would lead the researcher to reject both models, hence not providing any guidance to the researcher. The RV test instead leverages differences in model fit to help assess which model represents a better approximation of the actual supply side behavior.

References

Anderson, Theodore, and Herman Rubin. 1949. “"Estimation of the parameters of a single equation in a complete system of stochastic equations.” *Annals of Mathemetical Statistics*, 20(1): 46–63.

Backus, Matthew, Christopher Conlon, and Michael Sinkinson. 2020. “Common Ownership and Competition in the Ready-To-Eat Cereal Industry.” working paper.

Bergquist, Lauren Falcao, and Michael Dinerstein. 2019. “Competition and Entry in Agricultural Markets: Experimental Evidence from Kenya.” working paper.

Berry, Steven, and Philip Haile. 2014. “Identification in Differentiated Products Markets Using Market Level Data.” *Econometrica*, 82(5): 1749–1797.

Bonnet, Celine, and Pierre Dubois. 2010. “Inference on vertical contracts between manufacturers and retailers allowing for nonlinear pricing and resale price maintenance.” *RAND Journal of Economics*, 41(1): 139–164.

Bonnet, Celine, Pierre Dubois, Sofia Villas Boas, and Daniel Klapper. 2013. “Empirical Evidence on the Role of Wholesale Pricing and Vertical Restraints on Cost Pass-Through.” *Review of Economics and Statistics*, 95(2): 500–515.

Bresnahan, Timothy. 1982. “The Oligopoly Solution Concept is Identified.” *Economics Letters*, 10: 87–92.

Bresnahan, Timothy. 1987. “Competition and Collusion in the American Automobile Industry: The 1955 Price War.” *Journal of Industrial Economics*, 35(4): 457–482.

Byrne, David. 2015. “Testing Models of Differentiated Products Markets: Consolidation in the Cable TV Industry.” *International Economic Review*, 56(3): 805–850.

Ciliberto, Federico, and Jonathan Williams. 2014. “Does multimarket contact facilitate tacit collusion? Inference on conduct parameters in the airline industry.” *RAND Journal of Economics*, 45(4): 764–791.

Conlon, Christopher, and Jeff Gortmaker. 2019. “Best Practices for Differentiated Products Demand Estimation with pyblp.” working paper.

Crawford, Gregory, Robin Lee, Michael Whinston, and Ali Yurukoglu. 2019. “The Welfare Effects of Vertical Integration in Multichannel Television Markets.” *Econometrica*, 86(3): 891–954.

D’Haultfoeuille, Xavier, Isis Durrmeyer, and Philippe Fevrier. 2019. “Automobile Prices in Market Equilibrium with Unobserved Price Discrimination.” *Review of Economic Studies*, 86: 1973–1998.

Duarte, Marco, Lorenzo Magnolfi, and Camilla Roncoroni. 2020. "The Competitive Conduct of Consumer Cooperatives." working paper.

Eizenberg, Alon, and Dalia Shilian. 2019. "Structure, Conduct, and Contact: Competition in Closely-Related Markets." working paper.

Fan, Ying, and Christopher Sullivan. 2020. "Measuring Competition with a Flexible Supply Model." working paper.

Feenstra, Robert, and James Levinsohn. 1995. "Estimating Markups and Market Conduct with Multidimensional Product Attributes." *Review of Economic Studies*, 62: 19–52.

Gandhi, Amit, and Jean-Francois Houde. 2019. "Measuring Substitution Patterns in Differentiated Products Industries." working paper.

Gasmi, F, J Laffont, and Quang Vuong. 1992. "Econometric Analysis of Collusive Behavior in a Soft-Drink Market." *Journal of Economics and Management Strategy*, 1(2): 277–311.

Gayle, Philip. 2013. "On the Efficiency of Codeshare Contracts between Airlines: Is Double Marginalization Eliminated?" *American Economic Journal: Microeconomics*, 5(4): 244–273.

Genesove, David, and Wallace Mullin. 1998. "Testing Static Oligopoly Models: Conduct and Cost in the Sugar Industry, 1890-1914." *RAND Journal of Economics*, 29(2): 355–377.

Hall, Alastair, and Denis Pelletier. 2011. "Nonnested Testing in Models Estimated via Generalized Method of Moments." *Econometric Theory*, 27(2): 443–456.

Hansen, Bruce. 2020. *Econometrics.* unpublished.

Igami, Mitsuru, and Takuo Sugaya. 2019. "Measuring the Incentive to Collude: The Vitamin Cartels, 1990-1999." working paper.

Kennedy, Pauline, Daniel O'Brien, Minjae Song, and Keith Waehrer. 2017. "The Competitive Effects of Common Ownership: Economic Foundations and Empirical Evidence." working paper.

Lau, Lawrence. 1982. "On identifying the degree of competitiveness from industry price and output data." *Economics Letters*, 10(1-2): 93–99.

Magnolfi, Lorenzo, Daniel Quint, and Christopher Sullivan. 2020. "TBD." working paper.

Michel, Christian, and Stefan Weiergraeber. 2018. "Estimating Industry Conduct in Differentiated Products Markets." working paper.

Miller, Nathan, and Matthew Weinberg. 2017. "Understanding the Price Effects of the MillerCoors Joint Venture." *Econometrica*, 85(6): 1763–1791.

Nevo, Aviv. 2001. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica*, 69(2): 307–342.

Porter, Robert. 1983. "A Study of Cartel Stability: The Joint Executive Committee, 1880-1886." *Bell Journal of Economics*, 14(2): 301–314.

Rivers, Douglas, and Quang Vuong. 2002. "Model selection tests for nonlinear dynamic models." *Econometrics Journal*, 5: 1–39.

Schennach, Susanne, and Daniel Wilhelm. 2017. "A Simple Parametric Model Selection Test." *Journal of the American Statistical Association*, 112(520): 1663–1674.

Shi, Xiaoxia. 2015. "A Nondegenerate Vuong Test." *Quantitative Economics*, 6(1): 85–121.

Slade, Margaret. 1992. "Vancouver's Gasoline-Price Wars: An Empirical Exercise in Uncovering Supergame Strategies." *Review of Economic Studies*, 59(2): 257–276.

Sullivan, Christopher. 2020. "The Ice Cream Split: Empirically Distinguishing Price and Product Space Collusion." working paper.

Sullivan, Daniel. 1985. "Testing Hypotheses about Firm Behavior in the Cigarette Industry." *Journal of Political Economy*, 93(3): 586–598.

Varela, Mauricio, and Madhu Viswanathan. 2019. "Prices or Quantities?" working paper.

Villas-Boas, Sofia. 2007. "Vertical Relationships between Manufacturers and Retailers: Inference with Limited Data." *Review of Economic Studies*, 74(2): 625–652.

Vuong, Quang. 1989. "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrics Journal*, 57(2): 307–333.

A Proofs of Propositions

We first prove a useful lemma:

Lemma A.1. *Define the RSS measure of fit as*

$$Q_n^{RSS}(\theta,\gamma) = \frac{1}{n}\sum \hat{\omega}_{jt}(\theta,\gamma)^2. \tag{29}$$

Also define a GMM measure of fit constructed with the OLS moments as

$$Q_n^{OLS}(\theta,\gamma) = n g_n(\theta,\gamma)' W_n g_n(\theta,\gamma) \tag{30}$$

for

$$g_n(\theta,\gamma) = \frac{1}{n}[\Delta, w]'\hat{\omega}(\theta,\gamma), \tag{31}$$

and W_n equals to $([\Delta\ w]'[\Delta\ w])^{-1}$. Then, the two functions $Q_n^{RSS}(\theta,\gamma)$ and $Q_n^{OLS}(\theta,\gamma)$ are equal up to a constant, and have the same minimizers. Moreover, define the stacked vectors of regressors $\tilde{x} = [\Delta, w]$. Under the assumption that

$$plim(n^{-1}(p'p - p'\tilde{x}(\tilde{x}'\tilde{x})^{-1}\tilde{x}'p)) < \infty, \tag{32}$$

$plim(Q_n^{RSS}(\theta,\gamma))$ and $plim(Q_n^{OLS}(\theta,\gamma))$ are also equal up to a constant.

Proof of Lemma 1. Define the stacked vectors of parameters $\tilde{\theta} = [\theta,\gamma]$. We can thus rewrite (6) in matrix notation as:

$$n \cdot Q_n^{OLS}(\tilde{\theta}) = \hat{\omega}(\tilde{\theta})'\tilde{x}(\tilde{x}'\tilde{x})^{-1}\tilde{x}'\hat{\omega}(\tilde{\theta}) \tag{33}$$

$$= p'\tilde{x}(\tilde{x}'\tilde{x})^{-1}\tilde{x}'p - (\tilde{\theta}'\tilde{x}'\tilde{x}\tilde{\theta}) - 2\hat{\omega}(\tilde{\theta})'\tilde{x}\tilde{\theta} \tag{34}$$

$$= p'p - p'p + p'\tilde{x}(\tilde{x}'\tilde{x})^{-1}\tilde{x}'p - (\tilde{\theta}'\tilde{x}'\tilde{x}\tilde{\theta}) - 2\hat{\omega}(\tilde{\theta})'\tilde{x}\tilde{\theta} \tag{35}$$

$$= n \cdot Q_n^{RSS}(\tilde{\theta}) - (p'p - p'\tilde{x}(\tilde{x}'\tilde{x})^{-1}\tilde{x}'p) \tag{36}$$

$$= n \cdot Q_n^{RSS}(\tilde{\theta}) + c_n. \tag{37}$$

Since the functions Q_n^{OLS} and Q_n^{RSS} are equal up to the constant $\frac{c_n}{n}$, they also have the same minimizer $(\hat{\theta}_{OLS}, \hat{\gamma}_{OLS})$.

□

Proof of Proposition 1. Let $g_n(\theta,\gamma)$ be as defined in (31). Under the assumption that $\mathbb{E}[g_n(\theta,\gamma)] < \infty$, define

$$Q^{OLS}(\theta,\gamma) = \mathbb{E}[g_n(\theta,\gamma)]'W\mathbb{E}[g_n(\theta,\gamma)] \tag{38}$$

where W is a positive definite matrix, so that $Q_n^{OLS}(\theta,\gamma)$ converges uniformly to $Q^{OLS}(\theta,\gamma)$. Because W is positive definite and Q^{OLS} is a quadratic form, Q^{OLS} is a convex function of (θ,γ). By Lemma 1, $(\theta_{OLS},\gamma_{OLS})$ are the unique minimizers of Q^{OLS} and let – for two models $m = I, II$ – θ^m and $\hat{\gamma}^m$ be the values of θ specified by model m and the constrained GMM estimator of γ under model m when using the OLS moments.

We first establish that if $||\theta^I - \theta_{OLS}|| < ||\theta^{II} - \theta_{OLS}||$, then $||plim(\hat{\gamma}^I) - \gamma_{OLS}|| < ||plim(\hat{\gamma}^{II}) - \gamma_{OLS}||$. To see this, first consider that, for model m, using the definition of constrained GMM estimator we have

$$plim(\hat{\gamma}^m) = \gamma_{OLS} - plim(\frac{V_{21}}{V_{11}})(\theta_{OLS} - \theta^m),$$

where V_{ij} denotes elements (i,j) of the variance-covariance matrix of the GMM estimators.[8] Hence

$$\begin{aligned}||plim(\hat{\gamma}^I) - \gamma_{OLS}|| - ||plim(\hat{\gamma}^{II}) - \gamma_{OLS}|| &= ||plim(\frac{V_{21}}{V_{11}})(\theta^I - \theta_{OLS})|| - ||plim(\frac{V_{21}}{V_{11}})(\theta^{II} - \theta_{OLS})|| \\ &=\mid plim(\frac{V_{21}}{V_{11}}) \mid \Big(||\theta^I - \theta_{OLS}|| - ||\theta^{II} - \theta_{OLS}|| \Big).\end{aligned}$$

It follows that, if $||\theta^I - \theta_{OLS}|| < ||\theta^{II} - \theta_{OLS}||$, then

$$\left|\left| \begin{pmatrix} \theta^I \\ plim(\hat{\gamma}^I) \end{pmatrix} - \begin{pmatrix} \theta_{OLS} \\ \gamma_{OLS} \end{pmatrix} \right|\right| < \left|\left| \begin{pmatrix} \theta^{II} \\ plim(\hat{\gamma}^{II}) \end{pmatrix} - \begin{pmatrix} \theta_{OLS} \\ \gamma_{OLS} \end{pmatrix} \right|\right|,$$

and since the function Q^{OLS} is convex, then $plim(Q_n^{OLS}(\theta^I,\hat{\gamma}^I) - Q_n^{OLS}(\theta^{II},\hat{\gamma}^{II})) < 0$.

Finally, notice that by Lemma 1 the functions Q_n^{OLS} and Q_n^{RSS} are equal up to a constant, it follows that $plim(Q_n^{RSS,I} - Q_n^{RSS,II}) < 0$.

□

Proof of Proposition 2. Consider now the case where the parameter θ is estimated by minimizing the GMM criterion function Q_n as defined in Section 3. This function uses moments based on the vector of instruments (z, w). Hence, the corresponding estimator is consistent, so that the convex function $Q = plim(Q_n)$ has a unique minimum at the true parameter value (θ_0, γ_0). In the above proof of Proposition 1, one can then replace Q^{OLS} with Q and $(\theta_{OLS}, \gamma_{OLS})$ with (θ_0, γ_0), and the statement of Proposition 2 holds.

□

[8] For the definition of the restricted GMM estimator see e.g. Hansen (2020).

B Direction of the Bias in Estimating Conduct

Consider the special case of equation (3) where $\Delta^{II}_{jt} = 0$. This is the case where under model II firms price at marginal cost, similar as one of the hypotheses considered for consumer cooperatives by Duarte, Magnolfi and Roncoroni (2020). Then we have:

$$p_{jt} = \Delta_{jt}\theta + w_{jt}\gamma + \omega_{jt} \tag{39}$$

$$= \Delta^{I}_{jt}\theta_1 + c_{jt}. \tag{40}$$

In general, markups Δ_{jt} are functions of all market shares and prices, or $\Delta_{jt}(s_t, p_t)$, so that OLS estimates of equation (##) are biased. Because of the connection between testing based on RSS fit and OLS estimation of this equation, it is interesting to investigate the nature of the bias of the OLS estimate of θ.

In turn, analyzing the bias requires understanding how Δ_{jt} co-varies with ω_{jt} by considering the derivative $\frac{d\Delta_{jt}}{dc_{jt}}$ under a specific data generating process. By total differentiation

$$\frac{d\Delta_{jt}}{dc_{jt}} = \frac{\partial p_{jt}}{\partial c_{jt}} - 1, \tag{41}$$

so that the total derivative $\frac{d\Delta_{jt}}{dc_{jt}}$ equals the pass-through term $\frac{dp_{jt}}{dc_{jt}}$ minus one, and if pass-through is less than unit, there is negative mechanical co-variation between costs and markups. This would suggest that $cov(\Delta_{jt}, \omega_{jt}) < 0$, so that OLS estimates of θ are biased downwards.

However, we do not observe simple ceteris-paribus changes in costs in the data, and in particular there may be correlation between cost shocks ω_{jt} and demand shocks ξ_{jt}. Here we can distinguish two cases: if ξ_{jt} is a pure demand shock, the correlation with ω_{jt} may be negligible. If instead ξ_{jt} represents an unobserved product characteristic, then it is likely to be positively correlated with ω_{jt}. In the latter case, this will introduce a countervailing force to the downward bias we described above. In fact, ξ_{jt} is positively correlated to Δ_{jt} in many models of oligopoly competition such as Bertrand, and this may introduce a positive correlation between $cov(\Delta_{jt}$ and $\omega_{jt}) < 0$. This discussion highlights that the sign of the bias on θ is ambiguous.