Testing Firm Conduct

Marco Duarte* Lorenzo Magnolfi† Christopher Sullivan‡

November 1, 2020

Abstract

Understanding the nature of firm competition is of first order importance. In the empirical literature, researchers have used several different procedures to test two alternative models of firm conduct. In this paper, we formally compare the properties of these procedures, contrasting those designed for model assessment with those that perform model selection. In the absence of data on marginal costs, tests rely on a moment condition that must be formed using valid excluded instruments, e.g. demand rotators. We extend the intuition in Bresnahan (1982) and show that the testability condition in Berry and Haile (2014) requires - in a parametric environment - that the two candidate models generate different projections of markups on instruments. Without model misspecification, we show that a partial ranking in power exists among testing procedures in this setting. We then turn attention to the empirically relevant setting where either demand or cost is misspecified. Here we discuss the limitations of procedures designed to perform model assessment relative to model selection. With enough data, model assessment procedures will reject true models of conduct even for a small amount of demand or cost misspecification. However, the test developed in Rivers and Vuong (2002) may still be capable of selecting the true model.

*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.

1 Introduction

Firm conduct is both of independent interest for researchers (e.g. to detect collusive behavior, characterize conduct of non-profit firms, or determine the competitive effects of horizontal ownership), and a fundamental input into structural models of imperfectly competitive markets that are used to evaluate policy in many realms (e.g. competition enforcement, environmental regulation, trade policy). Often the true model of conduct is unknown. Thus, researchers typically perform econometric tests to evaluate a set of candidate models that are motivated by theory.

We consider the setting where theory suggests two alternative models of conduct. Examples of this setting in the industrial organization literature include investigations into the nature of vertical relationships, whether firms compete in prices or quantities, collusion versus competitive pricing, intra-firm internalization, common ownership, and nonprofit conduct. Papers in this area have used a variety of testing procedures, some of which are designed to perform *model assessment* and others for *model selection*. Model assessment independently pits each model against a set of alternatives, and uses the data to make a determination on whether that model can be rejected. Instead, model selection evaluates which of the two candidate models better fits the data. Although several approaches co-exist in the literature, to our knowledge there has been no systematic comparison of their properties.

In this paper we compare the econometric properties of four commonly used approaches to test firm conduct in a static model of supply and demand. These four approaches are: an estimation based (EB) procedure where testing is conducted as a by-product of estimation, a procedure that closely resembles an Anderson-Rubin test (AR), a Cox test, and a Rivers-Vuong test (RV). The first three are procedures designed for model assessment; the fourth is designed for model selection. We evaluate the properties of these procedures, considering both environments where all elements of the model are well-specified and also environments where some elements (such as demand or costs) may be misspecified.

We start our analysis in an ideal setting where the researcher has some information on either true marginal costs or markups, as it is the case in Nevo (2001). In this environment, different hypotheses on conduct can be evaluated by comparing the distributions of markups that they generate with the observed truth. This setting serves as a benchmark to illustrate the difference between model selection and model assessment.

More commonly, true markups are not observed, and we provide assumptions under which testing is possible in this case. We start from the intuition in Bresnahan (1982), which we extend to a standard setting with differentiated products and two hypotheses on conduct. We then formalize the connection between the nonparametric environment in Berry and Haile (2014) and this parametric setting. In doing so, we show that rejecting a wrong model of conduct in a large sample requires that model to generate different *predicted markups* (i.e. markups projected on the instruments) than the true model. This contrasts to the setting where true markups are observed and a model is falsified if it generates different markups than the true model. So when markups are unobserved, all of the four testing procedures that we examine need to rely on the same moment condition formed with excluded instruments. Motivated by examples in the literature, we show that conducting these tests without excluded instruments can yield misleading results.

Next, we consider these four tests in a setting where markups are measured without misspecification. All tests are consistent and have the right size under these conditions. Their power, however, is a function of the distance between the distributions of predicted markups for each candidate model and the true distribution of predicted markups. Using local power analysis, we are also able to provide a partial ranking between the four tests in this setting. Cox tests are the most powerful, and AR is weakly less powerful than EB. In fact, AR becomes increasingly less powerful as the number of instruments increases.

We then turn attention to the setting where either demand or marginal cost may be misspecified. Importantly, we show that misspecification - a likely occurrence in most empirical applications - alters the properties of these tests. In this setting, tests designed to do model assessment reject the true model of conduct for a large enough sample size. We also show that the RV approach may allow the researcher to conclude in favor of the true model. We formalize that, in the presence of misspecification, the test statistics for the four tests depend on both bias and noise. Here, we define bias as the distance between the distribution of predicted markups with and without misspecification, for the true model of conduct. Noise arises from both the variance of unobservables and sample size. If bias is low and the sample size is large enough, RV is the only test which concludes in favor of the true model of conduct, making it robust to mild misspecification. In fact, if the other tests conclude in favor of any model, it is due to noise. Given the progress made on demand estimation, and the increasing availability of data, we believe that such low-bias environments are likely to be prevalent in many applications, making RV an appealing testing procedure.

In Monte Carlo simulations, we show the empirical relevance of our results. First, simulations indicate that not using excluded instruments may lead to incorrect results. Second, we show that all four tests have good properties without misspecification. Finally, we show that RV is able to conclude in favor of the correct model in a misspecified setting with low bias, whereas the other three testing procedures increasingly reject both models of conduct as the sample size gets larger.

Overall, our results suggest that a researcher may want to adopt a model selection perspective when testing for firm conduct. As model selection only establishes that a model provides a better approximation of the truth than an alternative model, the RV test may be viewed as providing weaker evidence than tests designed to perform model assessment. In particular, one may be concerned that though one model fits better than another, both are far from the truth in an absolute sense. While these are valid concerns, our results show that in the presence of misspecification, model selection is the only realistic goal. However, to address the stated concerns, we recommend that a researcher supplement the results from an RV approach with estimation exercises that speak to the absolute distance between each candidate model and the data. In the Monte Carlo section, we illustrate this two-pronged approach.

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), Smith (1992), Gasmi, Laffont and Vuong (1992), Genesove and Mullin (1998), Nevo (2001), Ciliberto and Williams (2014), Miller and Weinberg (2017), Miller, Sheu and Weinberg (2019), Bergquist and Dinerstein (2020), Sullivan (2020), Fan and Sullivan (2020), Eizenberg and Shilian (2020)). Other important applications include common ownership (Kennedy, O'Brien, Song, and Waehrer (2017), Backus, Conlon and Sinkinson (2020)), vertical conduct (e.g. Villas-Boas (2007), Bonnet and Dubois (2010), Bonnet, Dubois, Villas-Boas, and Klapper (2013), Gayle (2013)), price discrimination (D'Haultfoeuille, Durrmeyer and Fevrier (2019)), price versus quantity setting (Feenstra and Levinsohn (1995)), 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, Lee, Whinston, and Yurukoglu (2019). Such data are very helpful to assess firm conduct. While we discuss this case in Section 3, we are 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 testing of non-nested hypotheses. Important surveys of this area include Gourieroux and Monfort (1994) and Pesaran and Weeks (2001). The AR approach is developed originally in Anderson and Rubin (1949) and is a central tool in the literature on inference in the presence of weak instruments (for a recent survey see Andrews, Stock and Sun (2019)). The Cox approach was introduced by Cox (1961) and specialized to GMM estimation in Smith (1992). 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.[1] We build on the insights of the econometrics literature that performs inference in the presence of model misspecification and highlights the importance of model selection procedures when models are misspecified (White (1982); Maasoumi and Phillips (1982); Hall and Inoue (2003); Marmer and Otsu (2012)).

The paper proceeds as follows. Section 2 describes the environment – a general model of firm conduct. Section 3 discusses testing when true costs or markups are observed, and introduces the concepts of model assessment and model selection. Section 4 formalizes the conditions under which we can falsify a model when true costs and markups are unobserved and introduces four approaches to testing in this setting: EB, AR, Cox, and RV. In Section 5 we establish properties of these testing procedures with no misspecification. Section 6 examines the choice of moments and the choice of weights in constructing the four test statistics. In Section 7, we turn attention to case of misspecification and establish the properties of the four approaches in this case. Section 8 presents

[1] More recently, work has been done to improve the asymptotic properties of the test. This includes Shi (2015) and Schennach and Wilhelm (2017).

evidence from Monte Carlo simulations. Informed by the theoretical and simulation results, in Section 9, we provide practical considerations for testing firm conduct and conclude. Proofs to all propositions are found in Appendix A.

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 compactness, we use i as a generic index for product j in market t. We observe price $\mathbf{p}_i$, market share $\mathbf{s}_i$, a vector of product characteristics $\mathbf{x}_i$ that affects demand, and a vector of characteristics $\mathbf{w}_i$ that affects the product's marginal cost. For each variable $\mathbf{y}_i$, we denote as $\mathbf{y}$ the full vector of $\mathbf{y}_i$ across the n observations, and $\mathbf{y}_t$ as the subvector of values in market t. We assume that, for all markets t, the demand system that generates elasticities is $s_t = \mathscr{s}(\mathbf{p}_t, \mathbf{x}_t, \boldsymbol{\xi}_t)$, where $\boldsymbol{\xi}_t$ is the vector of unobserved product characteristics. The equilibrium in market t is characterized by a system of first order conditions arising from the firms' profit maximization problems:

$$\mathbf{p} = \boldsymbol{\Delta}^*_{\mathbf{0}} + \mathbf{c_0}, \tag{1}$$

where $\boldsymbol{\Delta}^*_{\mathbf{0}}$ is the true vector of markup terms and $\mathbf{c_0}$ is the true vector of marginal costs. Following the literature, we assume that marginal costs are a linear function of observable cost shifters $\mathbf{w}$, so that $\mathbf{c_0} = \mathbf{w}\tau + \boldsymbol{\omega}_{\mathbf{0}}$.[2]

Cost shifters $\mathbf{w}$ are assumed to be exogenous throughout, so that $E[\mathbf{w}_i \boldsymbol{\omega}_{0i}] = 0$. Moreover, the coefficient τ is a nuisance parameter, as it is not the primary object of interest in this article where we focus on testing hypotheses on conduct. Hence, to streamline notation, we find it convenient to define orthogonalized versions of the main variables with respect to $\mathbf{w}$. For each variable $\mathbf{y}$, we denote its orthogonalized version with respect to the vector of cost shifters $\mathbf{w}$ as $y = M_{\mathbf{w}}\mathbf{y}$, where $M_{\mathbf{w}} = I - P_{\mathbf{w}}$ is the residual maker matrix and $P_{\mathbf{w}} = \mathbf{w}(\mathbf{w}'\mathbf{w})^{-1}\mathbf{w}'$ is the projection matrix.[3] Thus a bolded variable $\mathbf{y}$ is not orthogonalized while an unbolded variable y is orthogonalized with respect to $\mathbf{w}$.

The researcher can formulate alternative models of conduct m, estimate the demand system, and obtain estimates of markups $\hat{\Delta}_m$ under each model m. There is a large literature devoted to the estimation of elasticities and markups. To make progress on testing assumptions on the supply side, we abstract away from the estimation step and we treat $\Delta_{mi} = plim_{n\to\infty}\hat{\Delta}_{mi}$ as data.[4] Irrespective of the specific functional form of markups Δ_{mt}, in oligopoly models this markup term is generally

[2]This assumption is not necessary for our testing procedure to be valid, and could be relaxed.

[3]Throughout the paper we use projection matrices and residual maker matrices, defined for a generic vector $\mathbf{y}$ as $P_{\mathbf{y}} = \mathbf{y}(\mathbf{y}'\mathbf{y})^{-1}\mathbf{y}'$ and $M_{\mathbf{y}} = I - P_{\mathbf{y}}$.

[4]To speak to the practical application of testing procedures, we describe in Appendix B how to adjust standard errors to allow for two-step estimation, and also present Monte Carlo evidence where demand is estimated in a first step.

a function of both prices p_t and market shares s_t. Hence (1) for an assumed model m becomes:

$$p_t = \Delta_{mt}(p_t, s_t) + \omega_{mt}. \tag{2}$$

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

$$\begin{aligned} p =& [\Delta_1, \Delta_2]\theta + \omega \qquad (3) \\ =& \Delta\theta + \omega \qquad (4) \end{aligned}$$

where Δ is a n by two vector, and θ is a two by one vector. In the spirit of Atkinson (1970), equation (3) constructs an artificial nesting model and the two models correspond to different values of the parameter θ:

$$m = 1 : \theta = [1, 0]' \qquad vs \qquad m = 2 : \theta = [0, 1]'. \tag{5}$$

This nesting specification is general, and depending on the choice of Δ allows us to test many hypotheses on conduct found in the literature. Canonical examples of such hypotheses include the nature of vertical relationships, whether firms compete in prices or quantities, collusion, intra-firm internalization, common ownership and nonprofit conduct.

We assume that the true vector of markups Δ_0^* can be written as

$$\Delta_0^* = \Delta^*\theta_0, \tag{6}$$

so that the parameter θ_0 selects the markups for true model of conduct, and Δ^* contains well specified markups corresponding to the two candidate models. Without loss of generality, we assume throughout the article that $\theta_0 = [1, 0]'$, so that model 1 is the true model of conduct. It is possible that the vector of markups Δ observed by the researcher differs from the vector Δ^*. For example, one could estimate elasticities with a misspecified model of demand. Depending on the nature of the misspecification of markups implied by the two candidate models, different testing environments are possible. We consider in the paper three testing environments:

TE1 *No Misspecification* - We assume that model 1 is the true model and markups are not misspecified, so that $\Delta = \Delta^*$

TE2 *Markup Misspecification* - We assume that model 1 is the true model but markups are misspecified, so that $\Delta \neq \Delta^*$

TE3 *Model Misspecification* - We assume that neither of the two models 1 and 2 is the true model

To illustrate the difference between the three testing environments, consider the case of testing Bertrand competition versus joint profit maximization. In TE1, Bertrand may be the true model and markups are measured without error for both Bertrand and joint profit maximization. In

TE2, demand elasticities are mismeasured so that the markups implied by Bertrand and joint profit maximization are misspecified. In TE3, the true model is Cournot, so that neither of the two models specified by the researcher is true.

Another important consideration for testing conduct is whether markups for the true model Δ^*_0 are observed. We consider first in Section 3 the case where true markups are observed. We discuss instead the case where Δ^*_0 is unobserved in Sections 4 to 7 of the paper. In this case, estimating or testing hypotheses based on equation (2) requires instruments for Δ_{mt}, as markups are generally correlated with ω_{mt}. Note that this comes from two channels: one is direct, since Δ_{mt} is a function of p_t, which is an explicit function of ω_{mt}. The other is due to the fact that Δ_{mt} is a function of shares s_t, which in turn depend on the unobservable ξ_t, which may be correlated with ω_{mt}. Bresnahan (1982) and Berry and Haile (2014) provide intuition on what variation permits the researcher to distinguish two modes of conduct, 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 the following:

Assumption 1. *We assume that z_i is a vector of d_z excluded instruments for Δ^*, or $E[z_i\omega_{0i}] = 0$.*

To test hypotheses on θ, we construct a moment function with the excluded instruments z as:

$$g(\theta) = \frac{1}{n}z'(p - \Delta\theta), \tag{7}$$

$$= \frac{1}{n}z'\omega(\theta), \tag{8}$$

so that, for each model m, the moment function takes the value $g_m = g(\theta_m)$. We also introduce the first stage projection coefficient

$$\Gamma = E[z_i z_i']^{-1}E[z_i\Delta_i'] \tag{9}$$

$$= A_{zz}^{-1}A_{z\Delta}, \tag{10}$$

so that we define the projection of markups for model m on z as:

$$\tilde{\Delta}_m = z\Gamma\theta_m, \tag{11}$$

which we refer to as *predicted markups*. Finally, the following assumption summarizes regularity conditions that we maintain throughout:

Assumption 2. *Assume that:*

- *The data $(p_i, \Delta_i, \Delta^*_i, z_i), i = 1, 2, ...$ are jointly iid;*
- *The random variables p, z, Δ and Δ^* have finite fourth moments;*
- *The matrices $A_{\Delta\Delta} = E[\Delta_i\Delta_i']$, $A_{\Delta^*\Delta^*} = E[\Delta^*_i\Delta^{*\prime}_i]$, $A_{zz} = E[z_i z_i']$ and $E[\omega_0\omega_0' \mid z] = \Omega$ are positive semidefinite.*

These assumptions are necessary for GMM and OLS estimators of the parameter θ to be well defined, and help to derive the limiting distributions of the test statistics we discuss in Section 5.

3 Nevo's Menu Approach and Hypothesis Formulation

We consider in this section the ideal testing environment in which the researcher observes not only markups implied by the two candidate models $\Delta = \Delta^*$, but also the true markups (or equivalently marginal costs) Δ_0^*. For instance, accounting data may be available as in Nevo (2001).[5] If the testing environment is TE1, the researcher can proceed by comparing directly the markups Δ_m implied by each model m with the data Δ_0^*. In fact, the researcher would only have to do this for one observation.

Suppose instead that the researcher observes the distribution of Δ_0^*. Testing can be done by constructing a measure of lack of fit between the true distribution of markups, and the distribution of markups implied by each model m. Denote the lack of fit as $\mathscr{F}_m$. For example, $\mathscr{F}_m$ could be the difference between the median values of Δ_m and Δ_0^*, or between quantiles of the distributions of Δ_m and Δ_0^* as in Nevo (2001). One could also construct measures that compare the whole distributions such as Kolmogorov-Smirnov statistics, or the (squared) distance between markups $E[(\Delta_{0i}^* - \Delta_{mi})^2]$.

In order to evaluate models using the statistics $\mathscr{F}_m$, the researcher may proceed in one of two directions: *model selection*, or *model assessment*.[6] Model selection adopts a perspective where both models are treated symmetrically, and the model which better fits the data asymptotically is selected. Instead, model assessment concerns pitting a model m against a set of alternatives, and determining whether there is enough information in the data to reject that model.

These two strategies result in a different formulation of null and alternative hypotheses. The model assessment approach specifies for each model m a null hypothesis:

$$H_{0,m}: \lim_{n\to\infty} \mathscr{F}_m = 0 \tag{12}$$

versus an alternative of the form:

$$H_{A,m}: \lim_{n\to\infty} \mathscr{F}_m \neq 0. \tag{13}$$

By considering each model as a separate null hypothesis, this procedure relies on an absolute as opposed to relative assessment of each model. Hence, depending on the distances between the true markups and the implied markups for each model, four cases may arise. Either both nulls are rejected, only one is rejected, or neither is rejected. The researcher can conclude in favor of one model only when one null is rejected and the other null is not rejected. The failure to reject both nulls is not very informative, and may indicate that both models imply similar markups (see also

[5] In practice, Nevo (2001) observes a noisy measure of Δ_0^*. As a benchmark, we abstract from measurement error in the markup data.

[6] See e.g. Pesaran and Weeks (2001) for a general discussion of these two strategies.

the discussion in Gourieroux and Monfort (1994)). The rejection of both null hypotheses indicates misspecification, but gives no indications to the researcher on how to proceed.

Instead, we can construct a model selection procedure by formulating a null of the form:

$$H_0: \lim_{n\to\infty} \mathcal{F}_1 - \mathcal{F}_2 = 0 \tag{14}$$

Relative to this null, we can define two alternative hypotheses corresponding to cases of asymptotically better fit of one of the two models:

$$H_{A,1}: \lim_{n\to\infty} \mathcal{F}_1 - \mathcal{F}_2 = -\infty \qquad \text{and} \qquad H_{A,2}: \lim_{n\to\infty} \mathcal{F}_1 - \mathcal{F}_2 = +\infty \tag{15}$$

A model m is thus selected by this procedure if the null is rejected in favor of the alternative that corresponds to better asymptotic fit of model m. Intuitively, as long as the two models $m = 1, 2$ are asymptotically different, this procedure will select the one that generates markups that are closer to the true markups. So, whereas model assessment separately considers the absolute fit of each model, model selection considers the relative fit of the two models.

Hypothesis formulation thus interacts in an important way with the testing environments described above. In TE1, where markups are well-specified, direct comparison of the observed Δ_0^* and markups Δ_m^* implied by model m allows a researcher to directly falsify m. In environments TE2 and TE3, where markups and or the two hypotheses on conduct are misspecified, a model assessment approach will lead to the asymptotic rejection of both candidate models. Conversely, a model selection approach may allow the researcher to select the model that is closest to the truth, thus choosing asymptotically the true model as long as misspecification is not too severe. We explore more in depth the interaction between hypothesis formulation and misspecification in the common setup where the true markups are not observed, which we discuss in the next sections of the paper.

4 Testing Without Cost Data

We noted in the section above that one can perform testing in environment TE1 by simply comparing Δ_0^* with the markups implied by the two models when Δ_0^* are observed. Assuming no misspecification, this comparison would permit the researcher to determine the correct model.

In contrast, we assume in this section that we do not observe true markups, but we observe valid instruments that shift and rotate the demand curve. To build intuition, we continue to maintain environment $TE1$ where markups for both models are correctly specified and model 1 is the true model.[7]

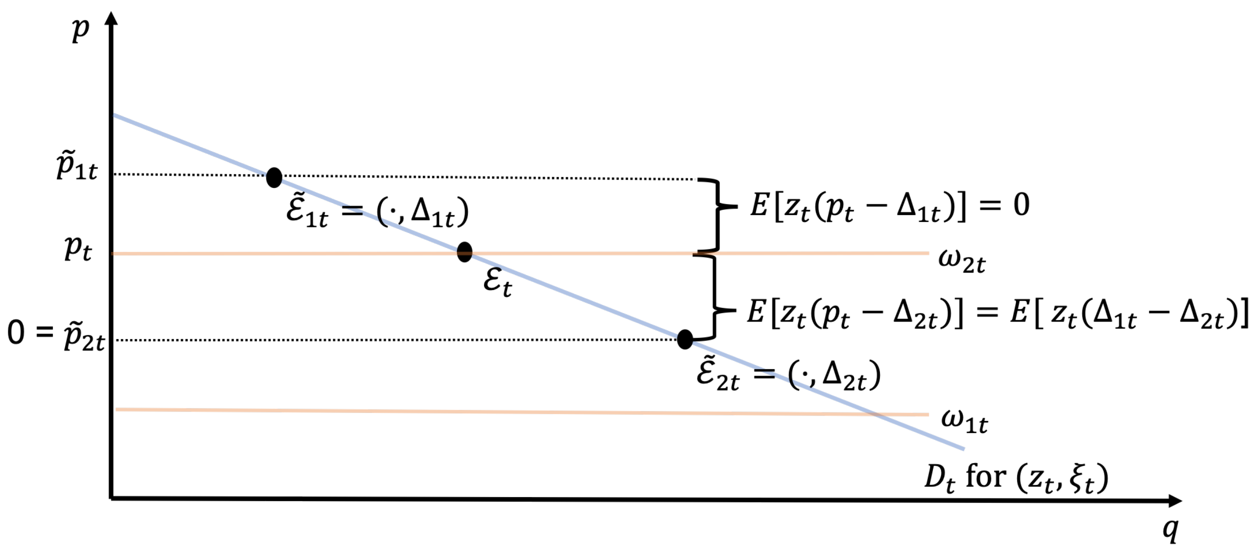
The canonical argument for distinguishing two models of conduct with this source of variation is laid out in Figure 2 of Bresnahan (1982), which illustrates how to distinguish the hypothesis of perfect competition from the one of monopoly. Suppose that we observe a market equilibrium $\mathcal{E}_{t'} = (\mathbf{p}_{t'}, \mathbf{q}_{t'})$ in period t'. With knowledge of the demand system, we can back out markups $\boldsymbol{\Delta}_{t'}$

[7] In later sections, we relax the conditions defined here for environments TE2 and TE3

corresponding to the two models of conduct, which imply corresponding marginal costs $(\mathbf{c}_{1t'}, \mathbf{c}_{2t'})$. Suppose that, in moving from period t' to period t, demand has rotated or shifted[8] while cost hasn't changed. Here we observe a new equilibrium $\mathcal{E}_t$. Since cost did not change, we can use the costs $(\mathbf{c}_{1t'}, \mathbf{c}_{2t'})$ to calculate the prices $(\tilde{\mathbf{p}}_{1t}, \tilde{\mathbf{p}}_{2t})$ in the new equilibrium predicted by the two models. We cannot falsify the model m for which $\mathbf{p}_t = \tilde{\mathbf{p}}_{mt}$.

While keeping costs fixed across different observations and markets may be theoretically possible, it does not correspond to what we observe in most empirical environments. In particular, costs depend on realizations of an unobservable ω, which vary across products and markets. Thus, we illustrate in Figure 1 the intuition behind testability of conduct. To do so, we extend Figure 2 in Bresnahan (1982) to our environment.

Figure 1: Testability of Conduct



Similar to Bresnahan (1982), Figure 2, we illustrate how to test hypotheses on conduct using instruments that shift and rotate the demand curve. In this case, model 1 corresponds to monopoly and model 2 corresponds to perfect competition.

In Figure 1, we orthogonalize all variables with respect to $\mathbf{w}$ - essentially keeping the observable part of marginal cost fixed across markets. For a large set of markets indexed by t, we observe the equilibrium $\mathcal{E}_t$, for which $p_t = \Delta^*_{0t} + \omega_{0t}$. However, we do not know the nature of conduct: following Bresnahan (1982) it could be monopoly, and the realized cost shock would have been ω_{1t}, or it could be perfect competition, and the corresponding cost shock would have been ω_{2t}. Across markets t, $\mathcal{E}_t$ will vary for three reasons: movement in demand induced by a vector of instruments z that satisfy Assumption 1, a new draw of the demand shock ξ, and a new draw of the cost shock ω_0.

In market t, we can use the demand system to solve for the implied markups Δ_{mt} under each model. Given that we have orthogonalized all variables with respect to w, $\tilde{p}_{mt} = \Delta_{mt}$ represents the predicted price under model m. We can then compare the observed prices with the prices predicted

[8]When marginal costs are constant, shifts of demand are sufficient for identification.

by each model. The predictions for both the true and the wrong model will differ from the observed prices. For the true model, this happens because of variation in the unobserved part of marginal cost ω_{0t}, which is uncorrelated with z_t. For the wrong model, there is an additional source of difference between realized and predicted prices arising from the different assumption on conduct. As the markups for each model are correlated with z_t, their difference may also be correlated with the instruments. This discussion provides a basis to falsify a model. We assume, as the foundation to testing conduct, that the wrong model is falsified:

Assumption 3. *The wrong model of conduct $m = 2$ is falsified, so that:*

$$E[z_i(p_i - \Delta^*_{2i})] \neq 0. \tag{16}$$

In TE1, $\Delta = \Delta^*$. Thus, given many markets, we cannot falsify a model m if $E[z_i(p_i - \Delta_{mi})] = 0$. Note that, given Equation (2), this condition is equivalent to $E[z_i\omega_{mi}] = 0$. Intuitively, this reflects the fact that our benchmark for whether a model is true is that the moment condition in Assumption 1 holds. This environment is a special case of the nonparametric testing setup described in Berry and Haile (2014).[9] In fact, after conditioning on $\mathbf{w}$, the marginal revenue function implied by any model m is $p_i - \Delta_{mi}$. So, under constant marginal costs, their condition (28) for a model to not be falsified becomes:

$$E[p_i - \Delta_{mi} \mid z_i] = 0 \qquad a.s., \tag{17}$$

which implies our condition above.

It useful to think about our ability to falsify models of conduct in terms of predicted markups as defined in Equation (11). Consider the setting in TE1 where model 1 is the true model: Equation (16) implies model 2 is falsifiable if $E[z_i\Delta_{1i}] \neq E[z_i\Delta_{2i}]$. This condition also ensures that the two models generally imply different vectors of predicted markups such that $\tilde{\Delta}_1 \neq \tilde{\Delta}_2$. This provides a parallel to the setting discussed in Section 3. There, given that true markups are observed, being able to falsify the incorrect model required the *level of markups* to differ for the two models, or $\Delta_1 \neq \Delta_2$. When true markups are unobserved, falsifiability of model 2 requires the *level of markups predicted by the instruments* to differ, or $\tilde{\Delta}_1 \neq \tilde{\Delta}_2$. The following lemma summarizes this discussion and presents a condition under which testing models is possible.

Lemma 1. *Suppose Assumption 1 holds, and let model $m = 1$ be the true model. Any other candidate model $m = 2$ is falsified if and only if:*

$$E[(\tilde{\Delta}_{1i} - \tilde{\Delta}_{2i})^2] \neq 0. \tag{18}$$

Thus, in TE1 condition (16) in Assumption 3 is equivalent to (18).

In light of the lemma, the ability to distinguish the two candidate models of conduct fundamentally requires having at least one instrument that generates different predicted markups for those models.

[9] See their Section 6, case (2).

For many pairs of hypotheses on firm conduct explored in applications, the condition in Equation (18) is typically satisfied. For instance, instruments always generate different predicted markups when the two candidate models imply strictly ranked markups, and for at least one of the two models the markups are a function of the instruments so that $E[z_i\Delta'_{mi}] > 0$.[10]

Although Lemma 1 provides a large sample condition under which testing is possible, it does not immediately suggest a testing procedure. To test alternative models of firm conduct in finite sample, we need to formulate hypotheses and construct test statistics. We discuss in the rest of this section the four main approaches taken in the literature. The first three: EB, AR, and Cox, are designed to perform model assessment, while the fourth, RV, adopts a model selection approach. Due to its popularity in the applied literature, we start with an estimation based approach where testing is conducted using Wald statistics.

4.1 Estimation Based Approach

Testing can be conducted as a by-product of estimation of the parameter θ.[11] We refer to this procedure as estimation based, or EB. 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). The estimating equation we consider is

$$p =\Delta\theta + \omega, \tag{21}$$

and for each model m we define null and alternative hypotheses:

$$H_0^{EB,m} : \text{plim}_{n\to\infty}\hat{\theta} = \theta_m \qquad vs \qquad H_A^{EB,m} : \text{plim}_{n\to\infty}\hat{\theta} \neq \theta_m, \tag{22}$$

where $\hat{\theta}$ is the GMM estimator of θ. The test can be implemented with the Wald approach by checking whether each model m implies parameters that are close to the estimates.[12] The corresponding

[10]In general, if instruments are shifters and rotators of demand, markups for both models will be functions of the instruments. An important exception is if one of the two models is perfect competition, whereby markups are identically zero. In this case the condition in the lemma is satisfied as long as for the alternative model m we have $E[z_i\Delta^{*\prime}_{mi}] > 0$.

[11]Although our formulation of the estimating equation is general in the sense that it can handle the comparison of any two models of conduct, it is not the only one. In some environments, such as the internalization matrix of example 2, the two models can be expressed as a nonlinear function of parameters κ, so that estimating equation is instead:

$$p = \Delta(\kappa) + \omega, \tag{19}$$

and the researcher may test hypotheses on the κ parameter. Alternatively, the researcher could formulate the model using a scalar parameter $\tilde{\theta}$ as:

$$p = \Delta_1\tilde{\theta} + \Delta_2(1-\tilde{\theta}) + \omega. \tag{20}$$

The main intuitions and results about the EB procedure in this paper apply to these formulations as well.

[12]Alternatively, the test could be implemented with the Likelihood Ratio approach by performing a J-test. As long as the GMM estimator is constructed using the optimal weight matrix, these approaches are equivalent. For ease of comparison with RV and AR, we focus our exposition on the Wald approach.

test statistic is thus:

$$T^{EB,m} = (\hat{\theta} - \theta_m)' \hat{V}_{\hat{\theta}}^{-1} (\hat{\theta} - \theta_m), \tag{23}$$

where $\hat{V}_{\hat{\theta}}$ is a consistent estimator of the variance of $\hat{\theta}$.

Notice that the condition in Lemma 1, while sufficient for falsifying model 2, is not sufficient for identification of the parameter θ, which is necessary for the EB procedure. When performing EB, we have then to maintain a stronger assumption:

Assumption 4a. *The matrix $A_{\Delta^* z} = E[\Delta_i^* z_i']$ is full rank.*

In particular, this assumption requires that we have at least two instruments in our setting.[13] This assumption implies the condition in Lemma 1.

As common in the literature, in our implementation of the test, we choose $\hat{\theta}$ to be the 2SLS estimator. We further discuss the choice of weight matrix in Section 6.2. Under the null hypothesis corresponding to model m, the statistic $T^{EB,m}$ is distributed according to a χ^2_2 distribution. Hence, comparing the value of $T^{EB,m}$ with the corresponding critical value from χ^2_2 yields a test with the correct size for the null corresponding to the true model m.

4.2 Anderson-Rubin Approach

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

$$p - \Delta_m = z\pi_m + \eta_m, \tag{24}$$

and then performs the test of the null hypothesis that $\pi = 0$ with a Wald test.[14] This procedure is similar to an Anderson and Rubin (1949) testing procedure, which accommodates hypothesis testing of endogenous linear models in the presence of weak instruments. For this reason, we refer to this procedure as AR. Recent examples of testing in this spirit are Backus, Conlon and Sinkinson (2020) and Bergquist and Dinerstein (2020). Formally, for each model m, we define the null and alternative hypotheses:

$$H_0^{AR,m} : \ \text{plim}_{n\to\infty} \hat{\pi}_m = 0, \qquad vs \qquad H_A^{AR,m} : \ \text{plim}_{n\to\infty} \hat{\pi}_m \neq 0. \tag{25}$$

where $\hat{\pi}_m$ is the OLS estimator of π under model m. We define the test statistic of the AR test for model m as:

$$T^{AR,m} = \hat{\pi}^{m\prime} \hat{V}_{\hat{\pi}_m}^{-1} \hat{\pi}_m \tag{26}$$

[13] The procedures to be introduced in Section 4.2-4.4, however, only require the condition in Lemma 1 to distinguish a wrong model from the true model. Thus, all three of these procedures could be implemented with only one instrument.

[14] The test could be implemented as a t-test if z is a scalar. The F-test and t-test implementations correspond to the Likelihood Ratio approach and Wald approach, respectively.

where $\hat{V}_{\hat{\pi}_m}$ is a consistent estimator of the variance of $\hat{\pi}_m$. If $\hat{V}_{\hat{\pi}_m}$ is the estimator under homoskedasticity, we have:

$$\hat{V}_{\hat{\pi}_m} = (z'z)^{-1}\hat{\sigma}_{\eta_m}, \tag{27}$$

where $\hat{\sigma}^2_{\eta_m}$ is an estimator of the variance of η_m. Thus our statistic takes on the form defined in Anderson and Rubin (1949) and Staiger and Stock (1997).[15] Under the null hypothesis corresponding to model m, the statistic $T^{AR,m}$ is distributed according to a $\chi^2_{d_z}$ distribution. Hence, comparing the value of $T^{AR,m}$ with the corresponding critical value from $\chi^2_{d_z}$ yields a test with the correct size for the null corresponding to the true model m.

4.3 Cox Approach

The next testing procedure we consider is inspired by the Cox (1961) approach to testing non-nested hypotheses. To perform a Cox test, we formulate two different pairs of null and alternative hypotheses for each model m, based on the moment conditions $g_m = g(\theta_m)$. Specifically, for model m we formulate:

$$H_0^{Cox,m} : \text{plim}_{n\to\infty} g_m = 0, \qquad vs \qquad H_A^{Cox,m} : \text{plim}_{n\to\infty} g_{-m} = 0, \tag{28}$$

where $-m$ denotes the alternative to model m. To implement the Cox test in our environment we follow Smith (1992). The test statistic we use for model m is:

$$T^{Cox,m} = \frac{\sqrt{n}}{\hat{\sigma}_{Cox}}\Big(g'_{-m}Wg_{-m} - g'_mWg_m - (g_{-m} - g_m)'W(g_{-m} - g_m)\Big), \tag{29}$$

where $\hat{\sigma}^2_{cox}$ is a consistent estimator of the asymptotic variance of the numerator of $T^{Cox,m}$ and W is a positive semidefinite weight matrix. As shown in Smith (1992), this statistic is distributed according to a standard normal distribution under the null hypothesis. The choices of moments and weights are further discussed in Section 6.

4.4 Rivers-Vuong approach

A prominent approach to testing non-nested hypotheses, which can be applied to testing firm conduct, is the one developed in Vuong (1989) and then extended to models defined by moment conditions in Rivers and Vuong (2002). Recent examples of this test in the applied literature are Molnar, Violi and Zhou (2013), Duarte, Magnolfi and Roncoroni (2020), and Sullivan (2020). Following Rivers and Vuong (2002), the null hypothesis for the test is that the two models $m = 1, 2$

[15] Although we use the variance estimator under homoscedasticity for analytical convenience in the derivations of this paper, it is easy to implement the test with the standard heteroscedasticity-consistent estimators of variance. Depending on the empirical context, the researcher may want to choose other forms of the estimator $\hat{V}_{\hat{\pi}_m}$ that allow, for instance, for clustered standard errors.

are $\sqrt{n}-$asymptotically equivalent, or

$$H_0^{RV}: \lim_{n\to\infty} \sqrt{n}\{Q_1 - Q_2\} = 0, \tag{30}$$

where Q_m is a measure of fit for model m. 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:

$$H_A^{RV,1}: \lim_{n\to\infty} \sqrt{n}\{Q_1 - Q_2\} = -\infty \qquad \text{and} \qquad H_A^{RV,2}: \lim_{n\to\infty} \sqrt{n}\{Q_1 - Q_2\} = +\infty \tag{31}$$

The test statistic is:

$$T^{RV} = \frac{\sqrt{n}}{\hat{\sigma}_{RV}}(Q_1 - Q_2), \tag{32}$$

where $\hat{\sigma}^2_{RV}$ is a consistent estimator of the asymptotic variance of the difference in the measures of fit. Rivers and Vuong (2002) show that the distribution of the test statistic T^{RV} is standard normal under the null.

In this paper 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. Given the sample moments defined in (7):

$$Q_m = g_m' W g_m \tag{33}$$

where W is a positive semi-definite weighting matrix. We describe the choice of moments and weights in greater detail in Section 6.

Having described the four testing procedures we consider in the article, we develop a comparison of their properties in the next section.

5 Comparison among tests without misspecification

Suppose that the researcher faces a testing environment of the type TE1 where true markups Δ_0^* are not observed, but well-specified markups $\Delta = \Delta^*$ are available. Suppose moreover that there exists a valid set of instruments as stated in Assumption 1, and that we use such instruments to form moments $g(\theta)$ as defined in Equation (7). As is well known, all four testing procedures defined in Section 4 - namely, EB, AR, Cox, and RV - have the correct size under these conditions.

In this section, we seek to gain further insight into the properties of these tests by comparing their power in the absence of misspecification. We first compare their power against fixed alternatives. To facilitate this comparison, we introduce the assumption of homoskedastic errors:

Assumption 5. *Assume that the error term ω_{0i} is iid, so that $E[\omega_{0i}^2|z] = \sigma^2_{\omega_0}$.*

For the RV and Cox testing procedures, which rely on GMM objective functions, we form the test statistics using the weight matrix $W = A_{zz}^{-1}$, as it is efficient under Assumption 5. We can then prove the following proposition:

Proposition 1. *Suppose that the testing environment is TE1, the true markups are unobserved, and that Assumptions 1, 2, 4a and 5 are satisfied. Then we can derive, for each test statistic, the following convergence in probability results:*

$$\text{(i)} \qquad \frac{T^{EB,m}}{n} \rightarrow_p \frac{E[(\tilde{\Delta}_{0i} - \tilde{\Delta}_{mi})^2]}{\sigma^2_{\omega_0}}$$

$$\text{(ii)} \qquad \frac{T^{AR,m}}{n} \rightarrow_p \frac{E[(\tilde{\Delta}_{0i} - \tilde{\Delta}_{mi})^2]}{\sigma^2_{\eta_m}}$$

$$\text{(iii)} \qquad \frac{T^{Cox,m}}{\sqrt{n}} \rightarrow_p \frac{2E[(\tilde{\Delta}_{0i} - \tilde{\Delta}_{mi})(\tilde{\Delta}_{-mi} - \tilde{\Delta}_{mi})]}{\sigma_{Cox}}$$

$$\text{(iv)} \qquad \frac{T^{RV}}{\sqrt{n}} \rightarrow_p \frac{E[(\tilde{\Delta}_{0i} - \tilde{\Delta}_{1i})^2 - (\tilde{\Delta}_{0i} - \tilde{\Delta}_{2i})^2]}{\sigma_{RV}},$$

so that all four testing procedures are consistent for $\theta_m \neq \theta_0$.

This result has immediate implications for the power of the tests with respect to two fixed alternative models $m = 1, 2$. For a test statistic T, we define the power of the test as:[16]

$$P[H_0 \text{ is rejected when } H_A \text{ is true}] = P[T > c_\alpha] \tag{34}$$

Hence, the proposition implies that - for fixed alternatives corresponding to the two models - all the tests we consider have asymptotic power of one.

More importantly, the proposition sheds light on the determinants of power in the tests we consider. Proposition 1 shows that the statistical power of these procedures depends on $E[(\tilde{\Delta}_{0i} - \tilde{\Delta}_{mi})^2]$, the distance between the *predicted* markups for the true model of conduct and the *predicted* markups implied by model m. Although the testing procedures that we consider differ, they all share the fundamental feature of being able to distinguish different models of conduct insofar as these models imply different predicted markups. Thus, in environment TE1, statistical power depends on the condition in Lemma 1 being satisfied, which requires having excluded instruments satisfying the moment condition in Assumption 1.

We can further compare the power of the different testing procedures by considering the case in which the two models of conduct represent local alternatives. This is a useful assumption often used in econometrics to describe alternatives that are "close" asymptotically. In our context, local alternatives are defined as follows:

Assumption 6a. *Assume that the models we consider represent local alternatives, or:*

$$\hat{A}_{z\Delta} = \frac{z'\Delta}{n} = \frac{\rho}{\sqrt{n}} + o_p(n^{-1/2}). \tag{35}$$

where ρ is a positive finite matrix.

[16] This formulation relies for simplicity on unilateral hypotheses. For RV, since there are two alternatives, this formulation captures the probability of rejection of the null in favor of the alternative corresponding to model $m = 2$.

Intuitively, this assumption restricts the covariation between instruments and markups implied by the two models to vanish as the sample size increases. In the spirit of local power analysis, this restriction serves as a tool to approximate the finite sample power of each test in an environment where the two models are hard to tell apart.[17] In this environment, models are characterized by rescaled versions of predicted markups, that are invariant to sample size. Hence we define:

$$\tilde{\Delta}^{(n)} = \sqrt{n}\tilde{\Delta} \tag{36}$$
$$= \rho' A_{zz}^{-1} z. \tag{37}$$

Using this assumption, we prove:

Proposition 2. *Assume that the testing environment is TE1, the true markups are unobserved, and that Assumptions 1, 2, 4a, 5 and 6a are satisfied. We also assume without loss that the true parameter is $\theta_0 = \theta_1$ and the null hypothesis for AR, EB and Cox is model $m = 2$. Then we can derive, for each test statistic, the following convergence in distribution results:*

$$\begin{aligned}
&(i) && T^{EB,m} \to_d \chi^2_2(c)\\
&(ii) && T^{AR,m} \to_d \chi^2_{d_z}(c)\\
&(iii) && (T^{Cox,m})^2 \to_d \chi^2_1(c)\\
&(iv) && (T^{RV})^2 \to_d \chi^2_1(c/4),
\end{aligned}$$

where we define the finite constant:

$$c = \frac{E\left[\left(\tilde{\Delta}^{(n)}_{0i} - \tilde{\Delta}^{(n)}_{mi}\right)^2\right]}{\sigma^2_{\omega_0}}. \tag{38}$$

This proposition makes it immediate to compute the power function for both the EB and AR tests in a local alternatives environment. To facilitate comparisons, we approximate power by using the results on the distribution of the square of the test statistic for RV and Cox. For RV, this means that power results from rejecting the null in favor of either alternative. However, as the noncentrality term c gets large, the probability of rejecting in favor of the wrong hypothesis becomes vanishingly small, so that power statements can be interpreted as probabilities of rejecting the null in favor of the right model. For Cox, the original test is one-sided, so that when considering the square of the test statistic, and deriving power based on its distribution, one can adjust the critical value to reflect the fact that the original test was one sided. To make the power of the Cox test directly comparable to the one of the alternative procedures, we consider instead a two-sided version of Cox, providing thus a lower bound on the actual asymptotic power of Cox.

Proposition 2 thus implies a definitive ranking in power in which Cox is the most powerful test, and EB is at least as powerful as AR. This ranking holds because the power of tests whose statistics

[17] We further comment on this assumption in Appendix A.4

are distributed as noncentral chi-squared with the same noncentrality depends on the number of degrees of freedom (DasGupta and Perlman (1974)), so that tests with fewer degrees of freedom have higher power. The power of RV relative to these three tests is ambiguous. While RV is always less powerful than Cox, its ranking with respect to AR and EB depends on the tradeoff between noncentrality and degrees of freedom.

It is interesting to trace the sources of these differences in power. Cox tests, because they compare the two models directly against each other, have the highest power in this setting. The gap in power between the EB and AR tests depends on the degree of overidentification as noted in the literature on weak instruments (e.g. Andrews, Stock and Sun (2019)). In settings where the number of instruments d_z is much larger than the dimension of the vector of markups $d_\Delta = 2$, as it is often the case in applications to the study of firm conduct, AR may be substantively less powerful than alternatives.[18][19]

6 Implementation of the tests

6.1 Choice of Moments

The tests described in Section 4 can all be written as functions of the moments defined in (7).[20] These moments were constructed using valid, excluded instruments z, so that the moment condition holds for the true model. The existence of such instruments is crucial for testing conduct, as discussed in Section 4. However, there exist implementations of the above tests in the literature which rely on other moment conditions to perform testing. Prominent examples are goodness of fit tests based on residual sum of squares (RSS). These tests have been popular in the applied literature in industrial organization.[21]

These RSS-based procedures are problematic and can provide misleading implications for conduct. To illustrate why, we briefly summarize how one could implement a test based on RSS in a model selection context. First, one would impose model m and rearrange the first order conditions of the firm as the following linear model in $\mathbf{w}$:

$$\mathbf{p} - \boldsymbol{\Delta}\theta_m = \mathbf{w}\tau + \boldsymbol{\omega}. \tag{39}$$

Under the imposed model, $\mathbf{p} - \boldsymbol{\Delta}\theta_m$ is observed and one can then obtain an OLS residual $\hat{\boldsymbol{\omega}}_m$ from the regression (39). Given that the OLS regression imposes that $\mathbf{w}$ is orthogonal to $\boldsymbol{\omega}_m$, one

[18] For instance, suppose that a researcher were to use BLP instruments or differentiation instruments to test conduct. These instruments are typically constructed so that there are 3 or more instrument per product characteristic. As noted in Gandhi and Houde (2019), this may result in a large set of instruments, especially when there are many product characteristics.

[19] We do not consider in this paper the consequences of weak instruments for testing. In this case, AR would seemingly offer advantages since its robustness to weak instruments is well understood in the econometrics literature.

[20] This is immediate from our definitions of RV and Cox, and easy to show for AR and EB.

[21] For example, Bresnahan (1987), Gasmi, Laffont and Vuong (1992), Villas-Boas (2007), Bonnet and Dubois (2010) and Bonnet, Dubois, Villas-Boas, and Klapper (2013) use RSS-based procedures to test hypotheses on conduct (collusion in the first two examples and vertical markets in the others).

can equivalently set $\hat{\boldsymbol{\omega}}_m = p - \Delta\theta_m = \hat{\omega}_m$. One then constructs for model m a measure of fit $Q_m^{RSS} = \frac{1}{n}\sum_i(\hat{\omega}_{mi})'(\hat{\omega}_{mi})$, and computes a test statistic analogous to (32):

$$T^{RSS} = \frac{\sqrt{n}}{\hat{\sigma}_{RSS}}(Q_1^{RSS} - Q_2^{RSS}), \tag{40}$$

where $\hat{\sigma}_{RSS}$ 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 4.4, except for the different measure of fit, making this procedure appear similar to RV. However, consider alternative moments

$$h_m = \frac{\Delta'\hat{\omega}_m}{n}, \tag{41}$$

which are the moments one would use to estimate Equation (3) by OLS. We define a measure of fit using these moments as $Q_m^{OLS} = h_m' W h_m$. Now we can prove the following lemma:

Lemma 2. *For any model m, Q_m^{RSS} and Q_m^{OLS} are equal up to a constant. Hence, the RSS test statistic can be written as:*

$$T^{RSS} = \frac{\sqrt{n}}{\hat{\sigma}_{RSS}}(Q_1^{OLS} - Q_2^{OLS}).$$

Due to the endogeneity of markups, the population analog of moments h_m will not be equal to zero at the true parameter value $\theta_m = \theta_0$. Instead, Q_m^{OLS} is minimized for $\theta_m = \hat{\theta}_{OLS}$, the OLS estimator of θ in Equation (3). Hence, testing based on T^{RSS} may generate tests with the wrong size, and possibly no power against some incorrect fixed alternatives. For example, consider the two models $\theta_1 = \theta_0, \theta_2 = \hat{\theta}_{OLS}$. As n goes to infinity, an RSS-based procedure using the statistic T^{RSS} will always reject the null of equal fit in favor of model 2, the wrong model. We formalize this intuition in the following proposition:

Proposition 3. *Assume that the testing environment is TE1, true markups are not observed, and that Assumptions 1, 2 and 5 are satisfied. Then we can derive the following convergence in probability results for T^{RSS}:*

$$\frac{T^{RSS}}{\sqrt{n}} \rightarrow_p \frac{E[(\Delta_{OLSi} - \Delta_{1i}))^2 - (\Delta_{OLSi} - \Delta_{2i}))^2]}{\sigma^{RSS}}, \tag{42}$$

where $\Delta_{OLS} = \Delta\theta_{OLS}$ and $\theta_{OLS} = E[\hat{\theta}_{OLS}] = A_{\Delta\Delta}^{-1}A_{\Delta p}$.

The proposition shows that the results of RSS-based tests are driven by the relative distance of the markups implied by the two models from the OLS predicted markups. Due to the fact that Δ is generally correlated with ω, the OLS estimates of θ are inconsistent, making the OLS projections of markups potentially far from the true markups, and thus skewing the test.

Our discussion thus far in this section is based on an RV-like test statistic because this has been used in the applied literature. In fact, other testing procedures (most notably, Cox and EB) could be performed using the moments h based on OLS fit, thus incurring in the same problems. An obvious solution is to test based on the correct moment g by including instruments as in the procedures described in Section 4. As we argue in Proposition 1, tests based on the moment condition g have the right size and are consistent for fixed alternatives, pointing asymptotically towards the model that generates predicted markups that are closest to the true ones.[22]

6.2 The Choice of Weights

An important aspect of the testing procedures we described is the choice of the weighting matrix. In RV and Cox, this choice involves the matrix W that enters in the GMM objective function. In EB, the choice of weights defines the estimator $\hat{\theta}$ that is used to construct the test statistic. In AR, although less apparent, a choice of weights is implicit in the use of the OLS estimator $\hat{\pi}_m$.

Hall and Pelletier (2011) highlight that the RV testing procedure is sensitive to the choice of weights. However, this is also true of Cox, EB, and AR - changing the weight matrix for a GMM estimator may have a substantial impact on parameter estimates and test statistics. Although the optimal weight matrix is efficient, using an approximation to it can work poorly in finite sample (e.g. Arellano and Bond (1991)) or when models are misspecified (Hall and Inoue (2003)), as in the testing environments TE2 and TE3 considered in the following section.

Thus, we adopt throughout the weight matrix $W = A_{zz}^{-1}$, which results in a number of desirable properties in our context. First, it is standard and optimal under homoskedasticity. Second, it facilitates comparison between different test statistics, generating expressions that share many similarities. Finally, it ensures that test statistics are simple function of the Euclidean distance between predicted markups, an intuitive object.

To see the latter two points, consider the general form for $T^{EB,m}$:

$$T^{EB,m} = (\hat{\theta} - \theta_m)'\hat{V}_{\hat{\theta}}^{-1}(\hat{\theta} - \theta_m) \tag{43}$$

$$= || \hat{\theta} - \theta_m ||^2_{\hat{V}_{\hat{\theta}}^{-1}} \tag{44}$$

In general, the EB test statistic can be interpreted as the distance between $\hat{\theta}$ and θ_m according to the distance function defined by the positive semidefinite matrix $\hat{V}_{\hat{\theta}}^{-1}$.

We showed above that the tests we consider depend on the Euclidean distance in predicted markups, which we related to the comparisons of markups performed in Nevo (2001), and to ideas of testability of models of conduct in Bresnahan (1982) and Berry and Haile (2014) through Lemma 1. We find it desirable to maintain this characterization of $T^{EB,m}$ as opposed to one depending

[22]In this paper, we have assumed Δ is data so that demand is known before testing supply. Practically, one could simultaneously estimate demand and supply jointly for the two candidate models and then do testing (see Gayle (2013) as an example). For this case, if one included enough demand instruments to overidentify the demand parameters by two, RSS measures of fit would satisfy the exclusion restrictions and be valid, so long as at least two of the demand instruments satisfy Assumptions 1 and 3.

on the distance of parameters according to the arbitrary norm defined by $\hat{V}_{\hat{\theta}}^{-1}$. Thus, under the assumption of homoskedasticity and for the choice of weight matrix $W = A_{zz}^{-1}$, the statistic $T^{EB,m}$ becomes:

$$T^{EB,m} = \frac{|| \tilde{\Delta}_0 - \tilde{\Delta}_m ||^2}{\hat{\sigma}^2_{\omega_0}} \tag{45}$$

Now, it is proportional to the squared Euclidean distance in predicted markups.

This characterization of the EB test statistic also enables easy comparisons with the other test statistics we consider. To see this point, note that the AR test statistic is in general equal to $T^{AR,m} =|| \hat{\pi}_m - \pi_0 ||^2_{\hat{V}_{\hat{\pi}}^{-1}}$. However, for the natural implementation of AR using the OLS estimator, we have:

$$T^{AR,m} = \frac{|| \tilde{\Delta}_0 - \tilde{\Delta}_m ||^2}{\hat{\sigma}^2_{\eta_m}}. \tag{46}$$

Moreover, under the simplifying assumption of homoskedasticity, it is immediate to show that:

$$nQ_m = \hat{\sigma}^2_{\omega_0} T_n^{EB,m}, \tag{47}$$

$$= \hat{\sigma}^2_{\eta_m} T_n^{AR,m}. \tag{48}$$

so that Q_m, used to compute both the RV and Cox test statistics, is equal to the test statistic for EB and AR up to a scale factor. Therefore, the assumption of homoskedasticity and the use of A_{zz}^{-1} as weight matrix ensure the test statistics depend on the same distance between predicted markups.

7 Comparison among tests with misspecification

As discussed in the previous section, under ideal conditions - namely, when the researcher has the correctly specified model of demand and valid instruments for both demand and supply - all four testing procedures are consistent. Practically, such an ideal environment is unlikely. We now turn attention to environment TE2 in which model 1 is the true model of conduct, but the markups for both models specified by the researcher are potentially misspecified. In this setting, markups could be misspecified due to either misspecification of demand or cost. We first consider the case of misspecification of the demand system. Then, we discuss the possibility that marginal costs, or both models of conduct themselves, are misspecified.

7.1 Testing with demand misspecification

We maintain here environment TE2, in which one of the two models of conduct we are considering is correct, but we allow markups to be misspecified through demand elasticities. In theory, demand systems could be estimated nonparametrically (e.g. Compiani (2020)), but data limitations or computational constraints currently preclude these approaches in most applications. Hence, researchers rely on parametric estimates of demand elasticities. While these can be very good approximations of reality, some misspecification is likely and can have implications for the testing procedures we describe.

In particular, demand misspecification introduces bias into our testing environment, and interesting differences arise in the relative performance of the procedures we consider. To formally discuss this kind of misspecification, we will distinguish between the true matrix of markups Δ^*, and the misspecified matrix of markups Δ which the researcher has available. To discuss testing in environment TE2 we introduce a modification of Assumption 4a:

Assumption 4b. *The matrix $A_{\Delta z} = E[\Delta_i z_i']$ is full rank.*

This assumption allows us to compute the EB test statistic with misspecified markups, and implies that the two models under consideration are testable even in an environment where markups are misspecified. To quantify the effect of misspecification, we define bias as the mean squared error in predicted markups under the true model of conduct:

$$B = E[(\tilde{\Delta}^*_{0i} - \tilde{\Delta}_{0i})^2]. \tag{49}$$

As long as bias is positive, it will be consequential for testing. Moreover, bias is most severe when it makes an incorrect model of firm conduct seem closer to the truth than the correct model of conduct. The two parts of the following assumption ensure that bias is present, but not severe:

Assumption 7. *We assume that TE2 is a low bias environment and model $m = 1$ is the true model of conduct, so that $B = E[(\tilde{\Delta}^*_{0i} - \tilde{\Delta}_{1i})^2]$, and:*

(i) $B > 0$, *and*

(ii) $B < E[(\tilde{\Delta}^*_{0i} - \tilde{\Delta}_{2i})^2]$.

Given the state of literature on demand estimation, and the increasing availability of more flexible and nonparametric methods, this assumption is sensible in many applied environments. We can then prove the following proposition:

Proposition 4. *Suppose that markups Δ^* are unobserved and that Assumptions 1, 2, 4a, 4b and 5 are satisfied. Then we can derive, for each test statistic, the following approximate convergence in probability results:*

$$(i) \qquad \frac{T^{EB,m}}{n} \rightarrow_p \frac{E\left[\left((\tilde{\tilde{\Delta}}^*_{0i} - \tilde{\Delta}_{0i}) + (\tilde{\Delta}_{0i} - \tilde{\Delta}_{mi})\right)^2\right]}{\sigma^2_{\omega_0}}$$

$$(ii) \qquad \frac{T^{AR,m}}{n} \rightarrow_p \frac{E\left[\left((\tilde{\Delta}^*_{0i} - \tilde{\Delta}_{0i}) + (\tilde{\Delta}_{0i} - \tilde{\Delta}_{mi})\right)^2\right]}{\sigma^2_{\omega_0}}$$

$$(iii) \qquad \frac{T^{Cox,m}}{\sqrt{n}} \rightarrow_p \frac{E[(\tilde{\Delta}^*_{0i} - \tilde{\Delta}_{-mi})^2] - E[(\tilde{\Delta}^*_{0i} - \tilde{\Delta}_{mi})^2] - E[(\tilde{\Delta}^*_{mi} - \tilde{\Delta}_{-mi})^2]}{2\sigma_{\omega_0}\sqrt{E[(\tilde{\Delta}_{1i} - \tilde{\Delta}_{2i})^2]}}$$

$$(iv) \qquad \frac{T^{RV}}{\sqrt{n}} \rightarrow_p \frac{E\left[\left((\tilde{\Delta}^*_{0i} - \tilde{\Delta}_{0i}) + (\tilde{\Delta}_{0i} - \tilde{\Delta}_{1i})\right)^2 - \left((\tilde{\Delta}^*_{0i} - \tilde{\Delta}_{0i}) + (\tilde{\Delta}_{0i} - \tilde{\Delta}_{2i})\right)^2\right]}{2\sigma_{\omega_0}\sqrt{E[(\tilde{\Delta}_{1i} - \tilde{\Delta}_{2i})^2]}},$$

where $\tilde{\tilde{\Delta}}^*_{0i} = \lambda'\tilde{\Delta}_i$ *and* $\lambda = E[\tilde{\Delta}_i\tilde{\Delta}_i']^{-1}E[\tilde{\Delta}_i\tilde{\Delta}^{*\prime}_{0i}]$.

The results in Proposition 4 have stark consequences on the asymptotic behavior of the testing procedures, which we highlight in the following Corollary:

Corollary 1. *Suppose that the testing environment is TE2 and satisfies the assumptions of Proposition 4. Then:*

1. *Under Assumption 7(i), RV rejects the null of equal fit, and the EB, AR and Cox testing procedures reject the null for $m = 1$ corresponding to the true model with probability that goes to one as the sample size increases;*

2. *Under Assumptions 7(i) and (ii), RV rejects the null of equal fit in favor of the alternative corresponding to the true model with probability that goes to one as the sample size increases.*

Corollary 1 highlights an important difference between RV and the other testing procedures. When markups are misspecified, EB, AR, and Cox (the model assessment procedures) reject the null corresponding to the correct model of conduct with probability that goes to one. Instead, RV (the model selection procedure) will reject the null of equal fit in favor of the true model as long as the misspecification is not too severe. In particular, RV will reject in favor of the true model $m = 1$ so long as it is closer to the truth in the sense of Assumption 7.

Note that Assumption 7 holds if and only if

$$E[(\tilde{\Delta}_{0i} - \tilde{\Delta}_{2i})^2] > -2E[(\tilde{\Delta}^*_{0i} - \tilde{\Delta}_{0i})(\tilde{\Delta}_{0i} - \tilde{\Delta}_{2i})]. \tag{50}$$

Intuitively, unless there is a sufficiently strong negative covariance between differences in predicted markups due to misspecification, and differences in predicted markups due to the wrong model of conduct, RV will reject in favor of the true model.

To discuss more formally the behavior of the testing procedures we analyze under misspecification of markups, we consider the limiting distribution of the test statistics under a true null in a local alternatives setting. The two forces at play here, which determine the probability of rejecting the null, are bias due to misspecification, and noise. We already defined bias as the term B, the mean squared error of markups corresponding to the true model. In general, we think of noise as determined by both $\sigma^2_{\omega_0}$, which captures the degree to which pricing is driven by random cost shocks, and the sample size n. Proposition 4 describes the probability limit of the test statistics as sample size gets larger, reducing noise. In the local alternatives environment the increase in precision due to larger samples is offset by the models getting closer to each other, thus shutting down sample size as a source of noise and allowing us to focus on the role of $\sigma^2_{\omega_0}$.

In the context of TE2, the local alternatives assumption needs to be augmented with a condition involving the true markups Δ^*:

Assumption 6b. *Assume that the models we consider represent local alternatives. In addition to Assumption 6a, also assume:*

$$\hat{A}_{z\Delta^*} - \hat{A}_{z\Delta} = \frac{\rho^*}{\sqrt{n}} + o_p(n^{-1/2}), \tag{51}$$

where ρ^* is a positive finite matrix.

We can thus establish the following:

Proposition 5. *Assume that the testing environment is TE2, that true markups are unobserved, and that assumptions 1, 2, 4a, 5, and 6b are satisfied. We also assume that the true parameter is $\theta_0 = \theta_1$ and the null hypothesis for AR, EB and Cox is model $m = 1$. Then we can derive, for each test statistic, the following convergence in distribution results:*

(i) $$T^{EB,1} \rightarrow_d \chi^2_2\left(\frac{B'^{(n)}}{\sigma^2_{\omega_0}}\right)$$

(ii) $$T^{AR,1} \rightarrow_d \chi^2_{d_z}\left(\frac{B^{(n)}}{\sigma^2_{\omega_0}}\right)$$

(iii) $$(T^{Cox,1})^2 \rightarrow_d \chi^2_1\left(\frac{(B^{(n)})^2}{4\sigma^2_{\omega_0}E[(\tilde{\Delta}^{(n)}_{1i} - \tilde{\Delta}^{(n)}_{2i})^2}\right)$$

(iv) $$(T^{RV})^2 \rightarrow_d \chi^2_1\left(\frac{(B^{(n)} - E[(\tilde{\Delta}^{*(n)}_{0i} - \tilde{\Delta}^{(n)}_{2i})^2])^2}{4\sigma^2_{\omega_0}E[(\tilde{\Delta}^{(n)}_{1i} - \tilde{\Delta}^{(n)}_{2i})^2]}\right)$$

where we define the modified bias constant:

$$B' = E[(\tilde{\tilde{\Delta}}^*_{0i} - \tilde{\Delta}_{0i})^2],$$

and all predicted markups and bias terms are computed using the stable $\tilde{\Delta}^{(n)}$.

The proposition shows that under misspecification the test statistics have noncentral χ^2 distributions whose noncentrality depends on bias and noise. For EB, AR and Cox, the noncentrality, and hence the probability of rejecting the true model of conduct, is proportional to bias (as measured by the stable versions of B and B') and inversely proportional to noise (as measured by $\sigma^2_{\omega_0}$). Practically, this implies that - when there is little noise in the testing environment - these procedures will reject true models of conduct even when the bias due to misspecification of markups is very small. While Proposition 4 shows that, in large samples, the null is rejected with probability that approaches one, Proposition 5 shows that, in finite samples, rejecting the null for these procedures depends on the relationship between bias and noise. Importantly, EB, AR, and Cox only fail to reject the true model of conduct when the noise is sufficiently large.

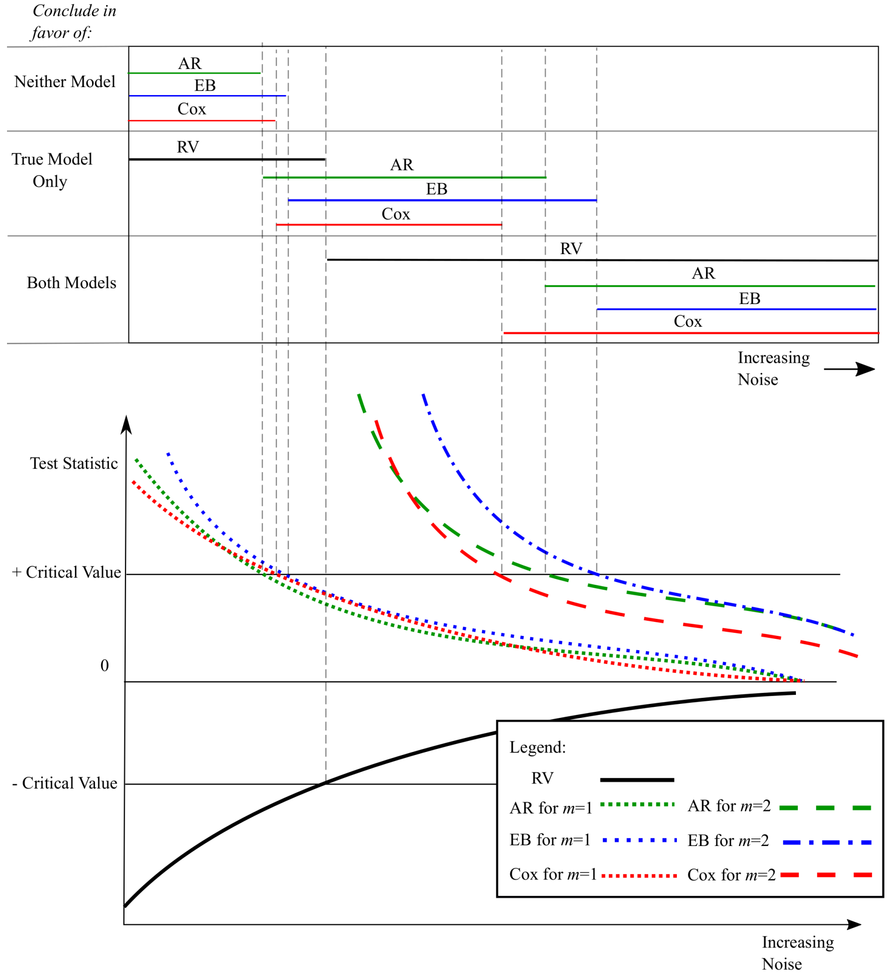
The RV test, on the other hand, behaves differently. The noncentrality, and thus the probability of rejecting the null of equal fit, depends on the quantity:

$$B^{(n)} - E[(\tilde{\Delta}^{*(n)}_{0i} - \tilde{\Delta}^{(n)}_{2i})^2], \tag{52}$$

the difference between two objects: the bias and a term equal to the distance between the predicted markups of the true model without misspecification and the predicted markups of the incorrect model $m = 2$ with misspecification. As long as the lack of fit of the incorrect model is larger than bias B as implied by Assumption 7, the RV test may reject the null in favor of the true model. Whether it does so depends again on noise. Crucially though, and unlike EB, AR, and Cox, less noise results in a higher probability of rejecting in favor of the true model for any given level of the difference in (52).

These points are illustrated in Figure 2. For a low bias environment, we plot the seven test statistics - these are the test statistic for RV, and statistics for the two different nulls for AR, EB and Cox - as a function of increasing noise. Intuitively, for all seven test statistics, values decline in magnitude as noise increases. As the test statistics cross critical values, what one concludes for each of the four tests changes. In addition to plotting test statistics, we present on the top panel schematics on how one would interpret the tests at each level of noise. We draw the figure under the assumption that model $m = 1$ is the true model.

Figure 2: Effect of Noise on Test Performance in a Misspecified Environment with Low-bias



This figure illustrates how the relative performance of the RV, AR, Cox and EB tests depends on noise. The top schematic describes how one would interpret the tests at each level of noise. The bottom graph plots the seven test statistics associated with the four tests against the critical values.

What drives the distinction between EB, AR and Cox on one side, and RV on the other, is how they formulate the null hypothesis. In turn, this difference is linked to the two fundamental goals of model assessment and model selection. When adopting a model assessment approach, we devise procedures that use the data to falsify models of conduct: this is what EB, AR and Cox test are designed to do. Crucially, any evidence that the predicted markups diverge from the true predicted markups will lead to rejection asymptotically. This is entirely appropriate when a researcher wants to learn whether a model generates correct predicted markups. However in the presence of misspecification, likely to be pervasive in practice, this approach has limitations. When the data become more informative, all models (even very close to be correct) will be rejected. Conversely, not rejecting often indicates that the data are not informative enough, rather than that the model is correct.

Alternatively, a researcher may pursue a model selection angle. While being open to the possibility of misspecification, we can learn which models of conduct better fits the data. This is what is accomplished with an RV procedure. In the model selection logic, pairwise comparisons between hypotheses are carried out examining their relative fit. As an analogy, model selection compares the relative fit of two candidate models and asks whether a "preponderance of the evidence" suggest that one model fits better than the other. Meanwhile, model assessment (and hence AR, EB, and Cox) uses a higher standard of evidence, asking whether a model is not falsifiable "beyond any reasonable doubt." While we may want to be able to conclude in favor of a model of conduct beyond any reasonable doubt, this is not a realistic goal in the presence of misspecification. However, if we lower the evidentiary standard, we can still make progress and learn about the true nature of firm conduct.

7.2 Misspecification of Marginal Cost

Above, we discussed the implications of demand misspecification on the testing procedures. Now, we turn our attention to the misspecification of marginal cost. Suppose:

$$\mathbf{p} = \boldsymbol{\Delta}^*_0 + \mathbf{w}\tau + \boldsymbol{\omega}_\mathbf{0}$$

is the DGP so that $E(\omega_\mathbf{0}|\mathbf{z},\mathbf{w}) = 0$ and the researcher observes Δ^*_0. However, the researcher uses misspecified cost-shifters $\mathbf{w_a}$ for testing conduct. The difference in $\mathbf{w}$ and $\mathbf{w_a}$ may be due to functional form (i.e. $\mathbf{w}$ contains higher order terms not contained in $\mathbf{w_a}$) or the omission/addition of relevant cost shifters. For each model m we can write:

$$\begin{aligned} \mathbf{p} &= [\boldsymbol{\Delta}^*_m + (\mathbf{w}\tau - \mathbf{w_a}\tau)] + \mathbf{w_a}\tau_\mathbf{a} + \boldsymbol{\omega}_m \\ &= \boldsymbol{\Delta}_m + \mathbf{w_a}\tau_\mathbf{a} + \boldsymbol{\omega}_m \end{aligned} \tag{53}$$

so that for any model m - including the true one - $\boldsymbol{\Delta}_m$ is different than $\boldsymbol{\Delta^*}_m$.

We assume that misspecification is consequential, but that the vector of cost shifters $\mathbf{w_a}$ only includes variables that are orthogonal to the original error term:

Assumption 8. *For the misspecified vector of cost shifters $\mathbf{w_a}$, the following holds:*

$$E[\boldsymbol{\omega}_\mathbf{0} \mid \mathbf{w_a}, \mathbf{z}] = 0. \tag{54}$$

This happens, for instance, if $\mathbf{w_a}$ is a subset of $\mathbf{w}$ and $\boldsymbol{\omega}_\mathbf{0}$ is mean independent of ($\mathbf{w}$,$\mathbf{z}$). Under Assumption 8, the exclusion restriction is satisfied for the true model of conduct with the wrong marginal cost included if markups are given as $\boldsymbol{\Delta}_\mathbf{0}$ in Equation (53). This transformation of the DGP shows that misspecification of costs can be turned into misspecification of markups.[23] Here

[23]If Assumption 8 does not hold, the misspecification of marginal cost is introducing an additional source of endogeneity, which is not addressed by the instrument vector z.

though, when a researcher uses misspecified costs $\mathbf{w_a}$, it is as if $\boldsymbol{\Delta}$ are the true unobserved markups, in the sense that they satisfy the DGP, while the researcher observes misspecified markups $\boldsymbol{\Delta}^*$.

Therefore, this shows that the effects of cost misspecification can be fully understood through the lens of markup misspecification. As we showed above, misspecification of markups which violates the moment condition affects all the testing procedures we have considered. And so, all the results derived in the previous section are relevant here. Thus, all the testing procedures have their power affected by misspecification of cost. However, while AR, EB, and Cox reject both models if given enough data, RV will select the right model asymptotically if the cost misspecification is not too severe.

7.3 Testing with model misspecification

In this subsection we consider the testing environment TE3, where neither of the two models is the true model and Δ_0^* is not observed. For instance, suppose that the true model in the DGP is Cournot competition, but the researcher decides to investigate whether firms compete a la Bertrand or pursue joint profit maximization.

In this environment, testing the hypotheses of Bertrand and joint profit maximization will not allow the researcher to learn the true mode of competition. A model assessment approach, at best, can lead to a rejection of both candidate models. However, if the ultimate goal is to learn which of the two models under consideration best approximates the truth, a model selection procedure can be useful. In particular, the researcher can still learn which model generates predictions of markups that are more accurate.

More formally, we first introduce the following assumption:

Assumption 9. *For any model of conduct $m = 1, 2$, the following inequality holds:*

$$E[(\tilde{\Delta}^*_{0i} - \tilde{\Delta}_{mi})^2] \neq 0 \tag{55}$$

This assumption, parallel to the condition in Lemma 1, ensures that the instruments allow the researcher to distinguish the markups associated with model m from the true markups. Given that neither model is correct in TE3, Assumption 9 relaxes the ranking on $E[(\tilde{\Delta}^*_{0i} - \tilde{\Delta}_{1i})^2]$ and $E[(\tilde{\Delta}^*_{0i} - \tilde{\Delta}_{2i})^2]$ imposed in Assumption 7. With Assumption 9, we can state the following Corollary to Proposition 4:

Corollary 2. *Suppose that the testing environment is TE3, and that Assumptions 1, 2, 4a, 4b, 5 and 9 are satisfied. Then:*

- *The EB, AR, and Cox testing procedures reject the null for any m with probability that goes to one as the sample size increases;*
- *RV rejects the null of equal fit in favor of the alternative m which is closer to the truth in the following sense:*

$$E[(\tilde{\Delta}^*_{0i} - \tilde{\Delta}_{mi})^2] < E[(\tilde{\Delta}^*_{0i} - \tilde{\Delta}_{-mi})^2].$$

In the testing environment TE3, where neither of the vectors of markups corresponding to the two candidate models is correct, the EB, AR, and Cox procedures will always reject any candidate model as long as the data are informative enough. Conversely, RV will point in the direction of the model that appears closer to the truth in terms of predicted markups. However, this relative assessment does not ensure that the chosen model m is close to the truth according to any absolute metric.

8 Monte Carlo Evidence

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

8.1 Setup

Simulation Model - 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 with characteristics $(\mathbf{x}_t, \mathbf{p}_t)$ and for which consumer i has an indirect utility of the form

$$U_{ijt} = \mathbf{x}_{jt}\beta + \mathbf{p}_{jt}\alpha + \boldsymbol{\xi}_{jt} + \boldsymbol{\mu}_{ijt} + \boldsymbol{\varepsilon}_{ijt} \tag{56}$$

where $\boldsymbol{\xi}_{jt}$ is an unobserved product-market specific shock to utility common to all consumers and $\boldsymbol{\mu}_{ijt}$ is an individual-specific taste for product attributes given as:

$$\mu_{ijt} = \mathbf{x}_{jt}\upsilon_{itx}\sigma_x + \mathbf{p}_{jt}\upsilon_{itp}\sigma_p.$$

where $\upsilon_{itx}, \upsilon_{itp} \sim N(0, 1)$. Finally, utility depends on an idiosyncratic shock ε_{ijt} which is distributed iid $T1EV$. The parameters governing consumer preferences are α, β and σ. For a given value of these parameters, the model implies a matrix $D_{\mathbf{p}_t} \equiv [\partial \mathbf{s}_j / \partial \mathbf{p}_t]$ of partial derivatives of product j's market share with respect to prices.

On the supply side we consider two candidate models of firm conduct that imply markups $\boldsymbol{\Delta}_m$ in Equation (2): Nash-Bertrand price setting (B) and a collusive model in which firms maximize joint profits (C). As in Section 2, costs are assumed to be the sum of a linear index of the observed cost shifters $\mathbf{w}$ and an unobserved component $\boldsymbol{\omega}$. For the two models we consider, markups are respectively:

$$\boldsymbol{\Delta}_{Bt} = (\Omega_{Bt} \odot D_{\mathbf{p}_t})^{-1}\mathbf{s}_{Bt} \tag{57}$$
$$\boldsymbol{\Delta}_{Ct} = (\Omega_{Ct} \odot D_{\mathbf{p}_t})^{-1}\mathbf{s}_{Ct}, \tag{58}$$

where Ω_{Bt} is the ownership matrix in market t where element (i,j) is an indicator that the same firm produces products i and j, and Ω_{Ct} is a $\mathcal{J}_t \times \mathcal{J}_t$ matrix of ones.

Parametrization - In our baseline, we have $N_m = 200$ markets, $N_f = 10$ firms and $N_p = 40$ products. We first allocate all products to all markets, and then randomly remove 40% of them, setting a higher probability of removing products from lower market indices. Firm-product pairs are randomly drawn to create an unbalanced ownership structure, where each firm can have at least 2 and at most 22 products. These allocation methods for product-market and firm-product create significant variation in market structure, which is important given the instruments that we use. The vector of product characteristics $\mathbf{x}$ contains a constant and a scalar $\mathbf{x}^{(1)}$. The vector of cost shifters $\mathbf{w}$ includes $\mathbf{x}$ and a scalar $\mathbf{w}^{(1)}$. Both $\mathbf{x}^{(1)}$ and $\mathbf{w}^{(1)}$ are independently drawn from a standard uniform distribution. In our baseline specification we set the following parameter values:

$$\begin{aligned} \text{Preference parameters:} &\quad \sigma_x = 3, \quad \sigma_p = 0.25, \quad \alpha = -1.5, \quad \beta = (-6, 6) \\ \text{Cost parameters:} &\quad \tau = (2, 1, 0.2) \\ \text{Unobserved shocks:} &\quad (\boldsymbol{\xi}_{jt}, \boldsymbol{\omega}_{jt}) \sim N(0, \Sigma), \quad \sigma^2_{\boldsymbol{\xi}} = \sigma^2_{\boldsymbol{\omega}} = 0.2 \text{ and } \sigma_{\boldsymbol{\xi}\boldsymbol{\omega}} = 0. \end{aligned}$$

Given parameter values, we draw observed $\mathbf{x}^{(1)}$, $\mathbf{w}^{(1)}$, and unobserved shocks for each of the $N_S = 100$ simulations and compute endogenous prices and quantities as the solution of the market share equations and first order conditions for each model m. The choice of parameter values and distributions create realistic outside shares for the Bertrand DGP, around 0.9.

Instruments - We construct two types of instruments for the Monte Carlo exercise: BLP instruments and differentiation instruments. For product-market jt, let O_{jt} be the set of products other than j sold by the firm that produces j and let R_{jt} be the set of products produced by rival firms. The instruments are:

$$\mathbf{z}^{BLP}_{jt} = \left[\sum_{k \in O_{jt}} 1[k \in O_{jt}] \quad \sum_{k \in R_{jt}} 1[k \in R_{jt}] \quad \sum_{k \in O_{jt}} \mathbf{x}_k \quad \sum_{k \in R_{jt}} \mathbf{x}_k\right]$$

$$\mathbf{z}^{Diff}_{jt} = \left[\sum_{k \in O_{jt}} \mathbf{d}^2_{jkt} \quad \sum_{k \in R_{jt}} \mathbf{d}^2_{jkt} \quad \sum_{k \in O_{jt}} 1[|\mathbf{d}_{jkt}| < sd(\mathbf{d})] \quad \sum_{k \in R_{jt}} 1[|\mathbf{d}_{jkt}| < sd(\mathbf{d})]\right]$$

where, following Gandhi and Houde (2019), $\mathbf{d}_{jkt} \equiv \mathbf{x}^{(1)}_{kt} - \mathbf{x}^{(1)}_{jt}$ and $sd(\mathbf{d})$ is the standard deviation across the distance vector. In most of the exercises, we use differentiation instruments and exogenous cost-shifters to estimate demand, and BLP instruments to perform supply side estimations.[24]

Implementation - We construct simulated data and compute demand estimates using the PyBLP package developed in Conlon and Gortmaker (2020). To solve for the equilibrium prices and shares as a fixed point, PyBLP uses the decomposition of the first order conditions in Morrow and Skerlos (2011). To compute demand estimates, PyBLP uses a nested fixed point algorithm and concentrates out the linear parameters from the two-step GMM minimization process. We use the default method in PyBLP for the fixed point (*simple*) and for the optimization (*l-bfgs-b*) routines, with stopping conditions *1e-12* and *gtol: 1e-5* respectively.[25] Integration is performed according to a level-9 Gauss Hermite product rule. Testing is performed using MATLAB. We account for the two-step estimation error in the computation of standard errors for all the testing procedures - see Appendix B for details.

8.2 Comparison between RV Test and RSS Test

In this subsection we illustrate the importance of constructing moments with valid instruments for testing. In particular, we show that endogeneity leads RSS-based tests to generate misleading conclusions. To emphasize this point, we abstract from demand estimation and use the true demand parameters to form markups for two models of firm behavior: Nash Bertrand and full collusion/joint profit maximization. In turn, we consider each of these two models as the DGP (in Panels A and B, respectively), and use both as candidate models for testing given the DGP. We choose different parameters (found in Table 1) to create simulation environments characterized by different amounts of endogeneity. Results for simulation environment (1) in Table 1 describe a case where endogeneity is small. For both DGPs, the RSS test and the RV test reject the null in favor of better fit of the correct model in every simulation.

Next we show that, RSS performs poorly as endogeneity increases. In statistical environment (2) we increase the amount of endogeneity by increasing the value of the demand parameter α, which determines consumers price sensitivity. The RV test still concludes in favor of the correct model in almost all simulations for both DGPs. Instead, in Panel A, the RSS test now concludes in favor of the correct Bertrand model in only 17 percent of simulations. In simulation environment (3), the problems with the RSS test become even more evident. Here we increase the endogeneity of markups by increasing the variance of $\boldsymbol{\omega}$, and observe that RSS rejects the null in favor of superior fit for joint profit maximization, the wrong model, in 79 percent of simulations. Simulation environment (4) adds additional endogeneity by increasing $\sigma_{\boldsymbol{\omega}\boldsymbol{\xi}}$, the covariance of the unobserved demand and cost shocks. While the performance of the RV test is largely unchanged, RSS now rejects the null

[24]In the environment of our simulations, we found that while differentiation instruments work better for the demand side, BLP instruments seem to perform better for supply side estimation. We compare instruments using Angrist-Pischke test statistic for weak instruments.

[25]We refer to PyBLP documentation website https://pyblp.readthedocs.io/en/stable/index.html for detailed information of the methods used.

Table 1: Test of Bertrand vs Joint Profit Maximization

		RSS Test			RV Test		
	Models	H_I: B	H_0	H_{II}: C	H_I: B	H_0	H_{II}: C
Panel A. DGP B	(1)	1.00	0.00	0.00	1.00	0.00	0.00
	(2)	0.17	0.80	0.03	1.00	0.00	0.00
	(3)	0.00	0.21	0.79	0.94	0.06	0.00
	(4)	0.00	0.10	0.90	0.93	0.07	0.00
Panel B. DGP C	(1)	0.00	0.00	1.00	0.00	0.00	1.00
	(2)	0.00	0.00	1.00	0.00	0.01	0.99
	(3)	0.00	0.00	1.00	0.00	0.13	0.87
	(4)	0.00	0.00	1.00	0.00	0.13	0.87

(1) $\sigma^2_\omega = 0.2,\ \sigma_{\xi\omega} = 0,\ \ \alpha = -1$

(2) $\sigma^2_\omega = 0.2,\ \sigma_{\xi\omega} = 0,\ \ \alpha = -1.5$ (baseline)

(3) $\sigma^2_\omega = 0.5,\ \sigma_{\xi\omega} = 0,\ \ \alpha = -1.5$

(4) $\sigma^2_\omega = 0.5,\ \sigma_{\xi\omega} = 0.3,\ \ \alpha = -1.5$

This table reports, for each simulation environment and for each hypothesis, the fraction of simulations in which the test either did not reject the null (for H_0) or concluded in favor of one alternative (for either H_m). Panels A and B correspond to Bertrand and joint profit maximization DGPs, respectively. Rows (1), (3) and (4) augment the baseline in the described way. Unless specified, parameters remain the same as in baseline.

in favor of the wrong model in 90 percent of the simulations, while never rejecting the null in favor of the correctly specified model. Taken together Panel A shows that in a simple simulation setting, the RSS test is highly susceptible to the degree of endogeneity of markups, and one cannot perform testing without valid instruments.

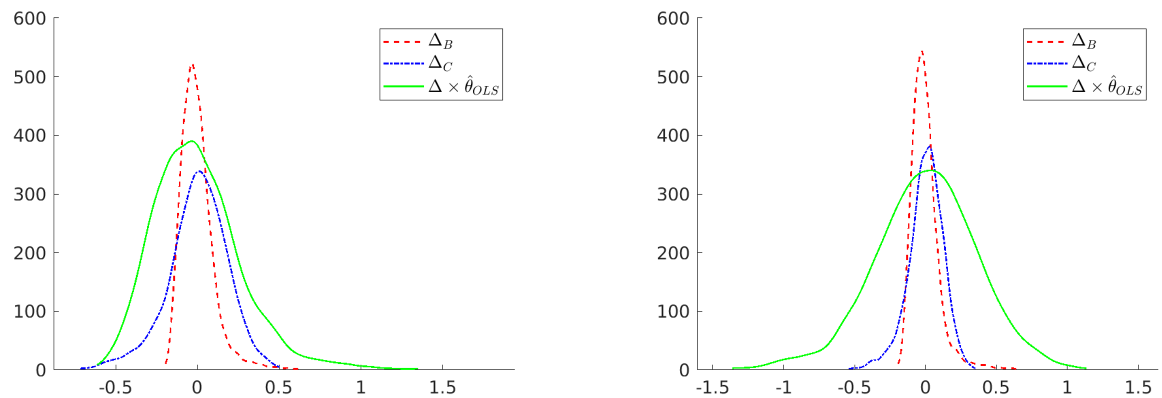
We then turn to analyze Panel B of Table 1, where the DGP is joint profit maximization. We can see that in all simulation environments, with either strong or weak endogeneity, the RSS test always rejects the null in favor of the correct model of joint profit maximization. It becomes apparent that in 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 Panel B, it does so for a reason that undermines the credibility of the test.

Proposition 3 allows us to rationalize the results in Table 1. RSS concludes in favor of the model whose implied markups are asymptotically closer to the markups implied by the OLS estimate of θ. In Figure 3 we plot the distribution of markups implied by the OLS estimate of θ. In both the low and high endogeneity simulation environments, the distributions of markups implied by the OLS estimate of the model are quite different from the distributions of markups implied by either the Bertrand or collusive models. In the low endogeneity environment (Panel A of Figure 3), markups implied by the OLS estimate happen to be closer to markups for the true rather than the wrong model. Thus, the RSS test is able to conclude in favor of the true model. As endogeneity increases (Panel B of Figure 3), the differences between the distribution of markups implied by the OLS estimate and either candidate model become large in an absolute sense. However, for both DGPs, the nature of the bias in this example is such that the distribution of markups implied by the OLS estimates is closer to the distribution for the collusive model, so the RSS test concludes in favor of the model of collusion.

FIGURE 3: Distribution of Markups implied by θ_{OLS}

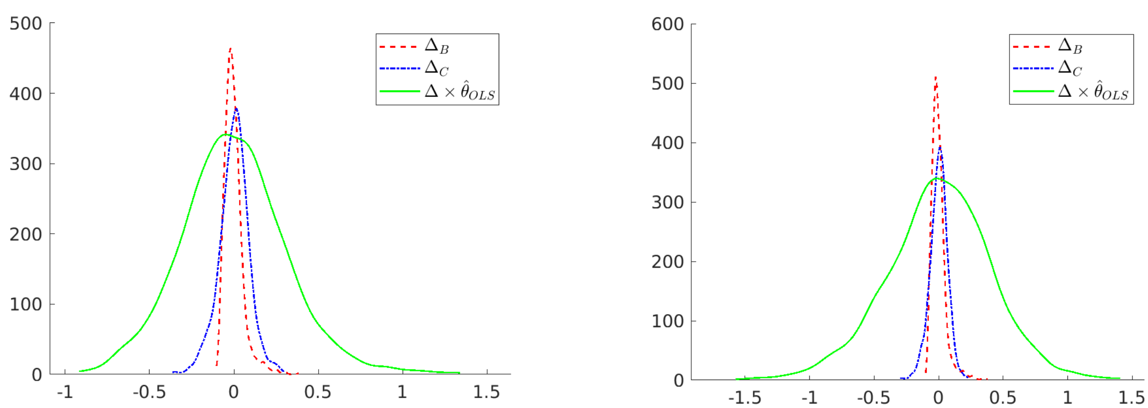
Panel A. Table 1 Environment (1) - Low Endogeneity

DGP = Bertrand | DGP = Collusion



Panel B. Table 1 Environment (4) - High Endogeneity

DGP = Bertrand | DGP = Collusion



This figure compares the distribution of predicted markups at the OLS estimate of θ to the predicted markups for the Bertrand and Collusive models. Panel A plots the predicted markups for simulation environment (1) in Table 1, a low endogeneity environment, whereas Panel B plots the distribution of predicted markups for simulation environment (4) in Table 1, a high endogeneity environment. In each panel, the left plot represents the distributions of markups when the true model is Bertrand and the right plot represents the distribution of markups when the true model is joint profit maximization.

8.3 Comparing performance between EB, AR, Cox and RV without misspecification

Given that we have shown that testing needs to be done with valid moments, we now turn to compare the four testing procedures when the tests are based on such moments. For a realistic comparison of the tests as they would be used in practice, we estimate demand in a first step to compute the markups implied by the two candidate models.[26] In Table 2, we present the results from comparing the performance of the four testing procedures in the baseline environment. We report in the table the fraction of simulations for which a specific test concludes in favor of an outcome.[27] Since all four approaches are designed to account for endogeneity, they perform well in a setting without any misspecification of demand. As we increase the sample size EB, AR, and Cox approach the correct size and RV always rejects the null in favor of the true model.

Table 2: Test Comparison - No Misspecification

		DGP B				DGP C			
	Models	RV	AR	EB	Cox	RV	AR	EB	Cox
Panel A. $N_m = 200$	B only	1.00	0.94	0.96	0.97	0.00	0.00	0.00	0.00
	C only	0.00	0.00	0.00	0.00	0.99	0.94	0.97	0.99
	$B = C$	0.00				0.01			
	Both		0.00	0.00	0.00		0.00	0.00	0.00
	Neither		0.06	0.04	0.03		0.06	0.03	0.01
Panel B. $N_m = 2000$	B only	1.00	0.92	0.93	0.97	0.00	0.00	0.00	0.00
	C only	0.00	0.00	0.00	0.00	1.00	0.94	0.95	0.96
	$B = C$	0.00				0.00			
	Both		0.00	0.00	0.00		0.00	0.00	0.00
	Neither		0.08	0.07	0.03		0.06	0.05	0.04

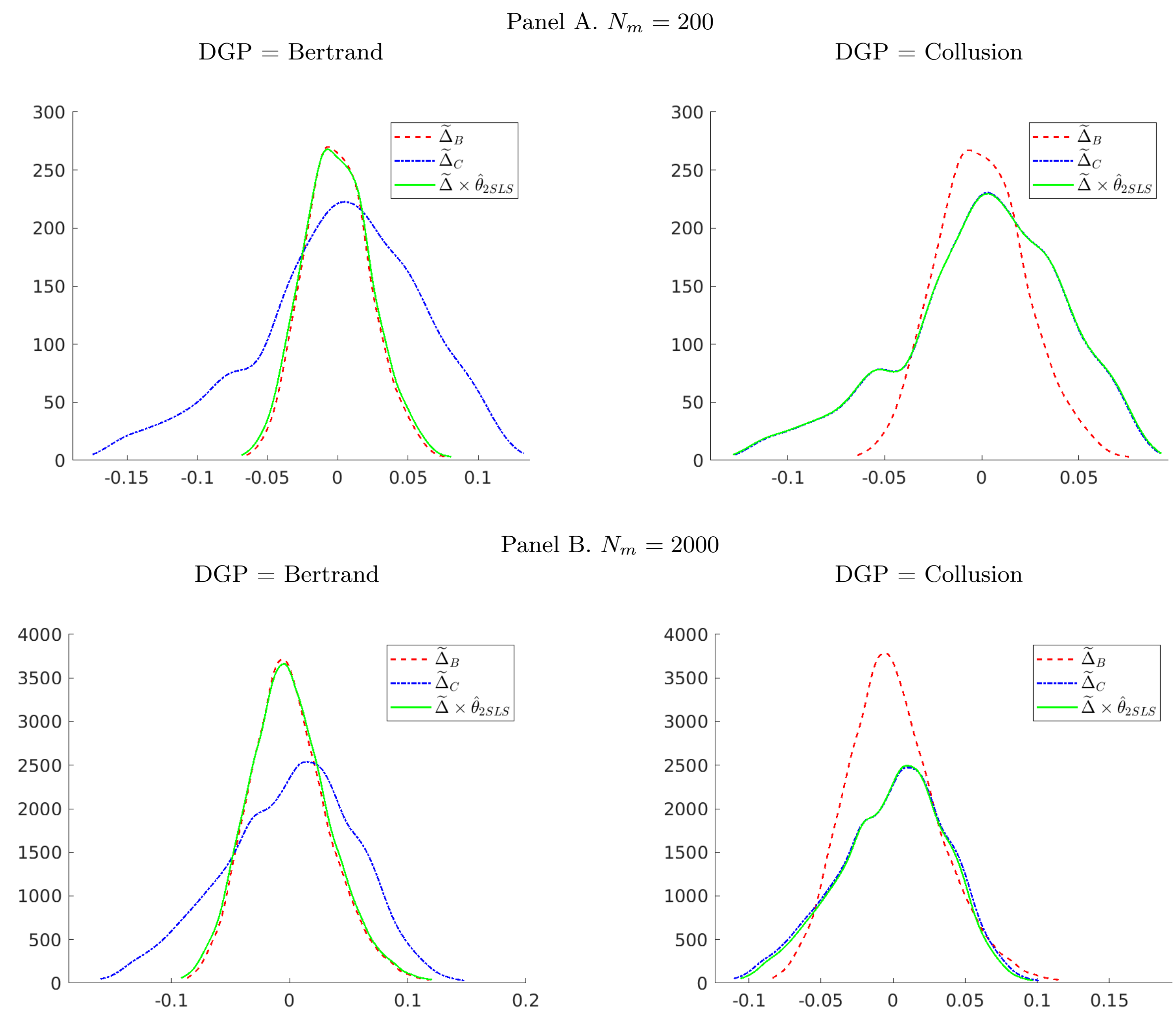
This table compares the performance of the four testing procedures: our RV procedure, the auxiliary regression based approach (AR), the estimation based approach (EB), and the Cox approach (Cox), for two models of conduct: Bertrand (DGP B) and joint profit maximization (DGP C). Rows correspond to all possible outcomes across these tests. Columns correspond to a particular testing environment applied to data generated by a specific DGP. Panels A and B report results for simulations where the number of markets N_m is set to 200 and 2000 respectively. We report in the table the fraction of simulations for which a specific test concludes in favor of a specific outcome.

[26] Hence, all the test statistics need to account for the error coming from the demand estimation. The details on how to compute standard errors in this two-step approach are in Appendix B.

[27] Notice that this statement has different meanings for different testing approaches. For instance, when RV concludes in favor of B only, it is rejecting the null of equal fit in favor of an alternative of better fit of B. Meanwhile for AR, EB and Cox concluding for B only means not rejecting the null that B is the true model while rejecting the null that C is the true model, pitted against all possible alternatives.

In Figure 4 we plot the distribution of predicted markups implied by the 2SLS estimate of θ in the setting with no misspecification. This figure illustrates why all four testing procedures work well in Table 2. When demand and costs are correctly specified, the predicted markups implied by the 2SLS estimates of θ very closely match the distribution of predicted markups for the true model. This approximation improves as the number of markets increases. As such, all four tests are consistent and have power against the wrong model of conduct. Thus, without misspecification, tests designed for model selection or model assessment conclude in favor of the true model in almost all simulations with either 200 (Panel A of Figure 4) or 2000 markets (Panel B of Figure 4).

FIGURE 4: Distribution of Markups implied by θ_{2SLS} with No Misspecification of Demand

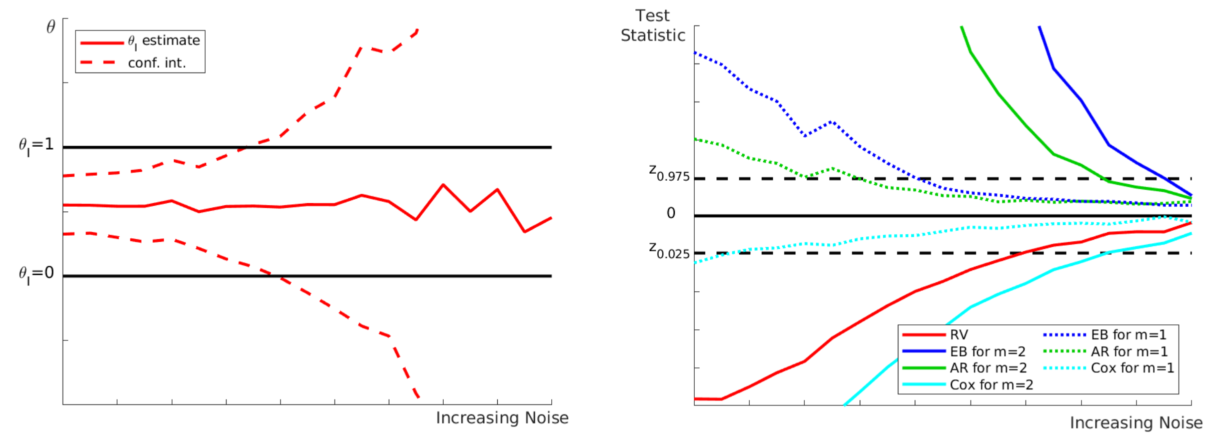


This figure compares the distribution of predicted markups at the 2SLS estimate of θ to the predicted markups for the Bertrand and Collusive models. Panel A plots the predicted markups estimated across 200 markets, whereas Panel B plots the distribution of predicted markups estimated across 2000 markets. In estimating demand and markups, the functional form of demand and cost are correctly specified. For each panel, the left plot represents the distributions of markups when the true model is Bertrand and the right plot represents the distribution of markups when the true model is joint profit maximization.

8.4 Comparing performance between EB, AR, Cox and RV with misspecification

We now consider the consequences of misspecification in TE2. We first explore the distinct forces of bias and noise illustrated in Figure 2 for the four tests. We construct a simple setup under a Bertrand DGP where demand (and hence markups under each model of supply) is not estimated, but bias is generated by altering one specific demand parameter value while holding others fixed at their true values.[28] This misspecification of the demand system generates incorrect markups, and hence biases the estimates of the supply-side parameter θ. For this given level of bias, we also vary the noise.[29] Figure 5 shows the results of our simulation.

FIGURE 5: Effect of Noise on Test Performance in Simulations



This figure illustrates how the relative performance of the RV, AR, EB and Cox tests depends on noise and bias in Monte Carlo simulations. The left graph shows estimates of the parameter θ_1 - the solid line plots the median estimate in 100 simulations, while the dashed lines plot the respective confidence intervals. Since we draw the figure under the assumption that model 1 is the true model, the true value of θ_1 is 1. The right graph plots the seven test statistics associated with the four tests against the critical values. Test statistics are normalized so that all share the same critical values.

In the left panel we plot the median estimate of the θ parameter that multiplies the joint profit maximization markups ($\hat{\theta}_1$) in the EB test setting. We also plot the median value of the confidence interval implied by each estimate. The bias created by the misspecification of demand is evident, and in a low noise environment it causes the true value of 1 to lie outside of the confidence interval for most of the simulations. As we add more noise to the estimation, confidence intervals become wider and eventually the true parameter lies inside the boundaries. However, if noise is large enough, then the value corresponding to the wrong hypothesis also lies inside the confidence interval. Thus noise has different implications for tests based model selection and tests based on hypotheses testing.

[28] In particular, while the true constant term in the indirect utility of consumers is -6, we alter it to -4.3.

[29] We do so by changing at the same time the number of markets N_m and the level of noise in the instruments. We alter the number of markets from 20 to 200 and alter the instruments so that $\mathbf{z} = \mathbf{z}^{BLP} + ((2 - n/50)^{1.6})\epsilon$ where n is the number of observations and ϵ is standard normal noise.

In the right panel, as in Figure 2, we plot the seven test statistics - these are the test statistic for RV, and statistics for the two different nulls of AR, EB, and Cox - as a function of increasing noise. All seven test statistics decline in absolute value as noise increases. For low noise, RV concludes in favor of the correct hypothesis while AR, EB and Cox reject both. As noise increases, RV becomes inconclusive, while AR, EB and Cox initially conclude for the correct hypothesis, but eventually do not reject either hypothesis. The conclusions one draws from the four tests under bias and noise are qualitatively the same as what was depicted in Figure 2. Thus, this simulation shows the empirical relevance of Proposition 5.

While Figure 5 illustrates the results proved in Section 7, it considers an environment where demand is not estimated and misspecification is introduced by directly choosing parameters of the demand system that depart from true values. To get closer to the challenges encountered by researchers, we conduct Monte Carlo simulations of test performance in a more realistic environments, where demand needs to be estimated and misspecification comes from a wrong choice of functional form for demand.

We introduce misspecification by estimating a logit demand system instead of a mixed logit. The misspecification of demand affects the markups used in the implementation of the tests and creates bias. In Table 3 we see that RV performs better than AR, EB and Cox for both DGPs and for both sample sizes. Compared with the results from Table 2, we see the differential impact of misspecification on the model selection versus the model assessment procedures. For a small sample size as in Panel A, the presence of noise leads the model assessment approaches to conclude in favor of the correct model in a quarter to a half of the simulations. However, if we increase the number of markets, as in Panel B, then EB, AR and Cox reject both alternatives in all simulations. In contrast, in a small number of markets RV rejects in favor of the true model in a vast majority of simulations, and as the number of markets increases, it does so in all simulations.

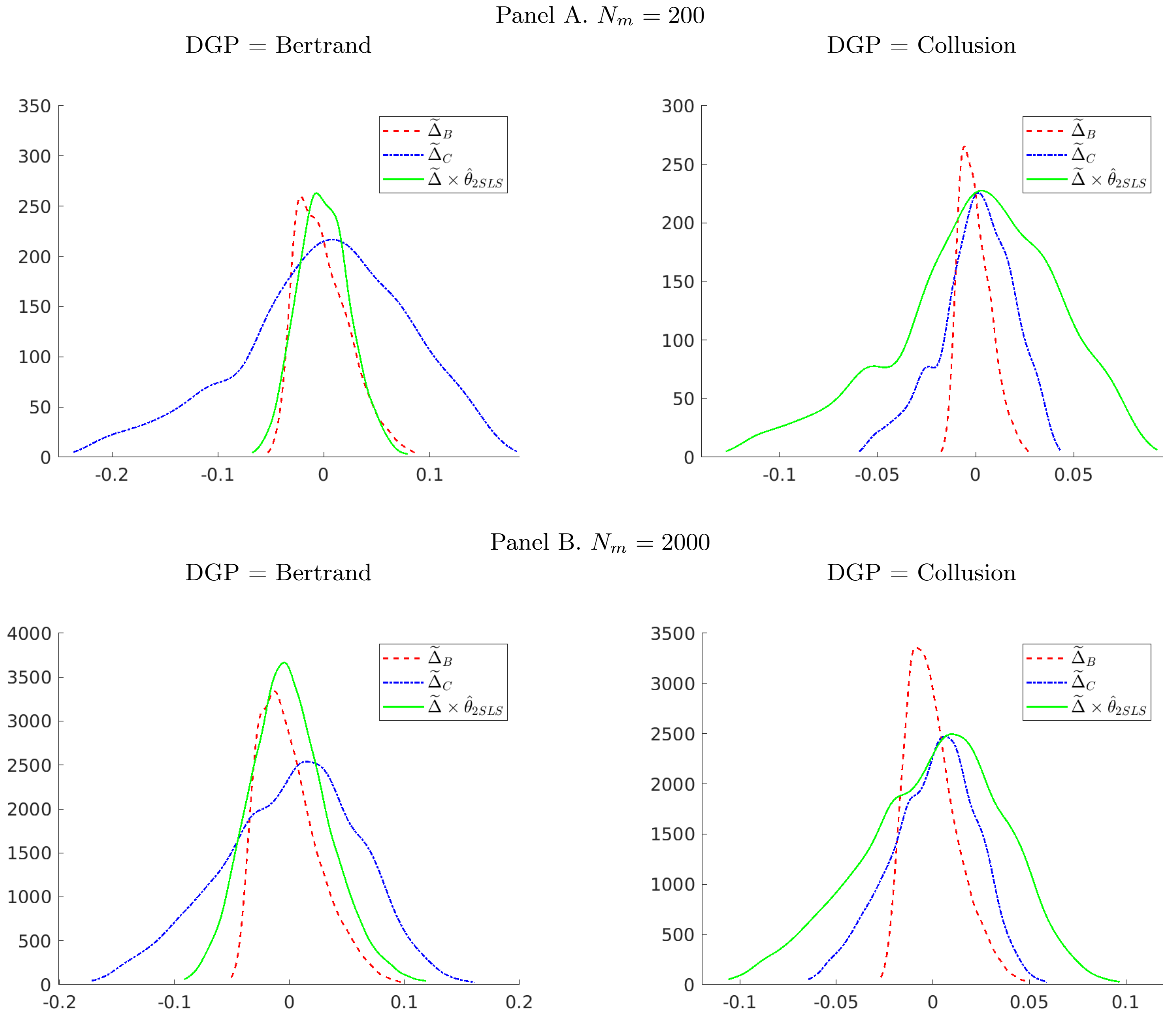
Table 3: Test Comparison - Misspecification

	Models	DGP B RV	DGP B AR	DGP B EB	DGP B Cox	DGP C RV	DGP C AR	DGP C EB	DGP C Cox
Panel A. $N_m = 200$	B only	0.78	0.48	0.41	0.24	0.00	0.00	0.00	0.00
	C only	0.00	0.00	0.00	0.07	0.98	0.43	0.39	0.37
	$B = C$	0.22				0.02			
	Both		0.01	0.00	0.17		0.00	0.00	0.02
	Neither		0.51	0.59	0.52		0.57	0.61	0.61
Panel B. $N_m = 2000$	B only	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	C only	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	$B = C$	0.00				0.00			
	Both		0.00	0.00	0.00		0.00	0.00	0.00
	Neither		1.00	1.00	1.00		1.00	1.00	1.00

This table compares the performance of the four testing procedures in an environment where demand is misspecified as logit. The true demand model in the DGP is a random coefficient model as in the baseline. Rows correspond to all possible outcomes across these tests. Columns correspond to a particular testing environment applied to data generated by a specific DGP. Panels A and B report results for simulations where the number of markets N_m is set to 200 and 2000 respectively. We report in the table the fraction of simulations for which a specific test concludes in favor of a specific outcome.

To better understand the source of these results, we plot, in Figure 6, the distribution of predicted markups implied by the 2SLS estimate of θ for one simulation. First, the figure shows that the distributions of predicted markups implied by the estimated model do not exactly fit the distribution of predicted markups implied by either candidate model. Further, this discrepancy is not reduced as the sample size increases. Thus, the tests designed for model assessment (EB, AR, and Cox) reject both models, as we see in the vast majority of simulations in Table 3. However, in this low bias environment, visual inspection suggests that the distribution of predicted markups implied by the estimated model is closer to the true model than the wrong model. As such, the RV test designed for model selection is able to conclude in favor of the correct model in every simulation for 2000 markets.

FIGURE 6: Distribution of Markups implied by θ_{2SLS} with Misspecification of Demand



This figure compares the distribution of predicted markups at the 2SLS estimate of θ to the predicted markups for the Bertrand and Collusive models. Panel A plots the predicted markups for estimation across 200 markets, whereas Panel B plots the distribution of predicted markups for 2000 markets. In estimating demand and markups, demand has been misspecified as logit. For each panel, the left plot represents the distributions of markups when the true model is Bertrand and the right plot represents the distribution of markups when the true model is joint profit maximization.

A final word on this simulation. Importantly, Propositions 4 and 5 reveal that in a low bias setting, RV can asymptotically conclude in favor of the true model of conduct. We believe that, given the advances in demand and the increasing availability of data, low bias environments are likely common. However, it is not feasible for a researcher to know the degree of the bias in their setting. In this simulation, we have misspecified demand in a way that might be considered severe, as a simple logit model likely generates substitution patterns that are less realistic than those produced by more flexible models used in practice. Yet we see that this is still a low bias environment, strengthening our belief that RV can be used in many settings to learn about firm conduct.

9 Conclusion and Practical Considerations for Testing

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 based 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, which is difficult to detect and is likely to arise in practice. Given our findings in Section 7 and our simulation results in Section 8, the model selection approach adopted in the RV test is the only one that can allow the researcher to learn the true model of conduct in the presence of misspecification.

A limitation of RV is it may conclude in favor of a model that, while relatively better in terms of asymptotic fit, may be far from the truth. Given the progress in estimating demand and costs, we believe that in practice researchers will find themselves often in a low-bias environment. However, it may still be advisable to present additional information beyond the results of RV testing. For instance, performing an estimation exercise in addition to RV testing can be highly informative. In fact, by recovering an estimate $\hat{\theta}$, one could obtain an estimate of predicted markups for the model implied by $\hat{\theta}$. We recommend comparing this estimated distribution of predicted markups to the distributions $\tilde{\Delta}_m$ implied by the two models. This step can show how far the models under consideration are from an estimate of the truth, thus giving a more complete picture of conduct. Hence, we suggest as best practice to report together with the RV results either $\hat{\theta}$, or the distribution of estimated predicted markups.

References

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

Andrews, Isaiah, James H Stock, and Liyang Sun. 2019. "Weak instruments in instrumental variables regression: Theory and practice." *Annual Review of Economics*, 11: 727–753.

Arellano, Manuel, and Stephen Bond. 1991. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations." *Review of Economic Studies*, 58(2): 277–297.

Atkinson, Anthony. 1970. "A method for discriminating between models." *Journal of the Royal Statistical Society*, 32(3): 323–345.

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. 2020. "Competition and Entry in Agricultural Markets: Experimental Evidence from Kenya." *American Economic Review*. forthcoming.

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.

Compiani, Giovanni. 2020. "Market Counterfactuals and the Specification of Multi-Product Demand: A Nonparametric Approach." working paper.

Conlon, Christopher, and Jeff Gortmaker. 2020. "Best Practices for Differentiated Products Demand Estimation with pyblp." working paper.

Cox, David. 1961. "Tests of separate families of hypotheses." *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, 1: 105–123.

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.

DasGupta, Somesh, and Michael Perlman. 1974. "Power of the Noncentral $F-$Test: Effect of Additional Variates on Hotelling's T^2-Test." *Journal of American Statistical Association*, 69: 174–180.

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. 2020. "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, Farid, Jean-Jacques 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.

Gourieroux, Christian, and Alain Monfort. 1994. "Testing non-nested hypotheses." *Handbook of Econometrics*, 4: 2583–2637.

Hall, Alastair, and Atsushi Inoue. 2003. "The large sample behaviour of the generalized method of moments estimator in misspecified models." *Journal of Econometrics*, 114(2): 361–394.

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

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.

Maasoumi, Esfandiar, and Peter Phillips. 1982. "On the behavior of inconsistent instrumental variable estimators." *Journal of Econometrics*, 19(2-3): 183–201.

Marmer, Vadim, and Taisuke Otsu. 2012. "Optimal comparison of misspecified moment restriction models under a chosen measure of fit." *Journal of Econometrics*, 170(2): 538–550.

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.

Miller, Nathan, Gloria Sheu, and Matthew Weinberg. 2019. "Oligopolistic price leadership and mergers: The united states beer industry." working paper.

Molnar, Jozsef, Roberto Violi, and Xiaolan Zhou. 2013. "Multimarket contact in Italian retail banking: Competition and welfare." *International Journal of Industrial Organization*, 31(5): 368–381.

Morrow, W. Ross, and Steven J. Skerlos. 2011. "Fixed-Point Approaches to Computing Bertrand-Nash Equilibrium Prices Under Mixed-Logit Demand." *Operations Research*, 59(2): 328–345.

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

Pesaran, M Hashem, and Melvyn Weeks. 2001. "Non-nested hypothesis testing: an overview." In *A Companion to Econometric Theory.* , ed. B. Baltagi, Chapter 13, 279–309. Oxford:Blackwell Publishers.

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.

Smith, Richard J. 1992. "Non-nested tests for competing models estimated by generalized method of moments." *Econometrica*, 973–980.

Staiger, Douglas, and James Stock. 1997. "Instrumental Variables with Weak Instruments." *Econometrica*, 65(3): 557–586.

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.

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.

White, Halbert. 1982. "Maximum likelihood estimation of misspecified models." *Econometrica*, 50(1): 1–25.

A Proofs

A.1 Preliminaries

We use throughout the notation:

$$\hat{A}_{zz} = \frac{z'z}{n}$$
$$\hat{A}_{\Delta z} = \frac{\Delta' z}{n}$$
$$\hat{A}_{\Delta^* z} = \frac{\Delta^{*\prime} z}{n}$$

When Assumptions 1 and 2 are true, the WLLN and CLT imply:

$$\hat{A}_{zz} = A_{zz} + o_p(1)$$
$$\hat{A}_{\Delta z} = A_{\Delta z} + o_p(1)$$
$$\hat{A}_{\Delta^* z} = A_{\Delta^* z} + o_p(1)$$
$$\frac{z'\omega_0}{\sqrt{n}} \rightarrow_d N(0, \Omega)$$

for $\Omega = E[z_i z_i' \omega_{0,i}^2]$.

We can define the (population) regression coefficient of endogenous markups on instruments as:

$$\Gamma^* = A_{zz}^{-1} A_{z\Delta^*} \tag{59}$$
$$\Gamma = A_{zz}^{-1} A_{z\Delta}. \tag{60}$$

Notice that:

$$A_{\Delta z} A_{zz}^{-1} A_{z\Delta} = A_{\Delta z} A_{zz}^{-1} A_{zz} A_{zz}^{-1} A_{z\Delta} \tag{61}$$
$$= \Gamma' E[z_i z_i'] \Gamma \tag{62}$$
$$= E[\Gamma' z_i z_i' \Gamma] \tag{63}$$
$$= E[\tilde{\Delta}_i \tilde{\Delta}_i'] \tag{64}$$

Suppose that the test statistic $T_n = a_n + b_n \frac{z'\omega_0}{n}$, where both $\sqrt{n} a_n$ and $\sqrt{n} b_n$ are sequences of random variables that converge in probability to constants. Under the assumption of homoskedasticity of ω_0, we define the asymptotic variance of $\sqrt{n}(T_n - plim(T_n))$ as follows:

$$Avar(\sqrt{n}(T_n) - plim(T_n)) = b_n' \Omega b_n \tag{65}$$
$$= \sigma_{\omega_0}^2 b_n' A_{zz} b_n. \tag{66}$$

A.2 Proof of Lemma 1

Proof: Let model 1 be the true model. We want to show that if and only if the condition

$$E[(\tilde{\Delta}_{1i} - \tilde{\Delta}_{2i})^2] \neq 0 \tag{67}$$

a candidate model $m = 2$ is falsified, in the sense of Definition 1:

$$E[z_i(p_i - \Delta_{2i})] \neq 0. \tag{68}$$

As a useful preliminary, recall that the first stage is:

$$\begin{aligned} \Delta_i &= z_i\Gamma + \eta_i \\ &= \tilde{\Delta}_i + \eta_i, \end{aligned}$$

where η is projection error orthogonal to z by construction, so that $E[z_i\eta_i] = 0$.

We first prove that (67) implies (68). Notice that a necessary and sufficient for (67) to be true is that there exists some nonzero measure subset Z of the support of z_i such that for $z_i \in Z$ we have $\tilde{\Delta}_{1i} \neq \tilde{\Delta}_{2i}$. In turn, because of the definition of projection, it must be that for $z_i \in Z$ we have $z_i\Gamma_1 \neq z_i\Gamma_2$. Then, we know that $\Gamma_1 \neq \Gamma_2$ which in turn implies $E[z_i\Delta_{1i}] \neq E[z_i\Delta_{2i}]$ since $\Gamma = E[z_iz_i']^{-1}E[z_i\Delta_i]$. Condition (68) follows immediately from the fact that model $m = 1$ is the true model.

Now we prove that (68) implies (67). To do so, we proceed by contradiction: suppose that (68) holds, but $E[(\tilde{\Delta}_{1i} - \tilde{\Delta}_{2i})^2] = 0$. Then, almost surely $\tilde{\Delta}_{1i} = \tilde{\Delta}_{2i}$. Hence we can write:

$$E[z_i(p_i - \tilde{\Delta}_{1i})] = E[z_i(p_i - \tilde{\Delta}_{2i})],$$

and since $E[z_i\eta_i] = 0$ we have:

$$E[z_i(p_i - \tilde{\Delta}_{1i} - \eta_{1i})] = E[z_i(p_i - \tilde{\Delta}_{2i} - \eta_{2i})],$$

which leads to a contradiction of (68). Hence, when (68) holds, (67) must hold as well.

A.3 Proof of Proposition 1

Proof: We start by proving the first part of the proposition in which we show that the modified test statistics converge in probability to constants that depend on $E[(\tilde{\Delta}_{0i} - \tilde{\Delta}_{mi})^2]$.

(i) - EB Consider the test statistic $T^{EB,m}$ under the assumptions of weight matrix $W = A_{zz}^{-1}$ and homoskedasticity. We first consider the object $\hat{\theta}$:

$$\hat{\theta} = \theta_0 + (\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\hat{A}_{z\Delta})^{-1}\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\frac{z'\omega_0}{n}$$

and

$$\hat{V}_{\hat{\theta}} = \frac{\hat{\sigma}^2_{\omega_0}}{n}(\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\hat{A}_{z\Delta})^{-1} \tag{69}$$

$$= \hat{\sigma}^2_{\omega_0}(\Delta' z(z'z)^{-1}z'\Delta)^{-1} \tag{70}$$

so that:

$$T^{EB,m} = (\theta_0 - \theta_m)'\hat{V}_{\hat{\theta}}^{-1}(\theta_0 - \theta_m) \tag{71}$$

$$+ \frac{\frac{z'\omega_0}{\sqrt{n}}\hat{A}_{zz}^{-1}\hat{A}_{z\Delta}(\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\hat{A}_{z\Delta})^{-1}\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\frac{z'\omega_0}{\sqrt{n}}}{\hat{\sigma}^2_{\omega_0}} \tag{72}$$

$$+ 2n\frac{(\theta_0 - \theta_m)'\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\frac{z'\omega_0}{n}}{\hat{\sigma}^2_{\omega_0}} \tag{73}$$

$$= \frac{(\tilde{\Delta}_0 - \tilde{\Delta}_m)'(\tilde{\Delta}_0 - \tilde{\Delta}_m)}{\hat{\sigma}^2_{\omega_0}} + \zeta_{EB}'\zeta_{EB} + \frac{2n}{\hat{\sigma}^2_{\omega_0}}\frac{(\tilde{\Delta}_0 - \tilde{\Delta}_m)'\omega_0}{n} \tag{74}$$

To derive the plim of the test statistic $T^{EB,m}$, we first consider the object $\hat{\theta}$:

$$\begin{aligned}
\hat{\theta} &= (\Delta' P_z \Delta)^{-1}\Delta' P_z(\Delta\theta_0 + \omega_0) \\
&= \theta_0 + (\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\hat{A}_{z\Delta})^{-1}\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\frac{z'\omega_0}{n} \\
&\to_p \theta_0.
\end{aligned}$$

Now, from standard GMM theory the asymptotic variance of $\hat{\theta}$ under homoskedasticity of the error term is

$$V_\theta = \sigma^2_{\omega_0}(A_{\Delta z}A_{zz}^{-1}A_{z\Delta})^{-1} \tag{75}$$

$$= \sigma^2_{\omega_0}E[\tilde{\Delta}_i\tilde{\Delta}_i']^{-1}. \tag{76}$$

Since $\hat{V}_\theta$ is a consistent estimator of V_θ we have:

$$\frac{T^{EB,m}}{n} \to_p \frac{(\theta_0 - \theta_m)'E[\tilde{\Delta}_i\tilde{\Delta}_i'](\theta_0 - \theta_m)}{\sigma^2_{\omega_0}} \tag{77}$$

$$= \frac{E[(\tilde{\Delta}_{0i} - \tilde{\Delta}_{mi})^2]}{\sigma^2_{\omega_0}} \tag{78}$$

(ii) - AR To derive the plim of the test statistic $T^{AR,m}$, we first consider the object $\hat{\pi}_m$

$$\begin{aligned}\hat{\pi}_m &= (z'z)^{-1}z'(\Delta\theta_0 + \omega_0 - \Delta\theta_m)\\ &= \hat{A}_{zz}^{-1}\hat{A}_{z'\Delta}(\theta_0 - \theta_m) + \hat{A}_{zz}^{-1}\frac{z'\omega_0}{n}\\ &\to_p A_{zz}^{-1}A_{z'\Delta}(\theta_0 - \theta_m).\end{aligned}$$

Under the assumption of homoskedasticity, V_{π_m} has the usual OLS form of:

$$V_{\pi_m} = \sigma^2_{\eta_m} A_{zz}^{-1} \tag{79}$$

Since $\hat{V}_{\pi_m}$ is constructed to be a consistent estimator of V_{π_m} we have

$$\frac{T^{AR,m}}{n} \to_p (\theta_0 - \theta_m)' A_{\Delta' z} A_{zz}^{-1} V_{\pi_m}^{-1} A_{zz}^{-1} A_{z'\Delta}(\theta_0 - \theta_m) \tag{80}$$

$$= \frac{(\theta_0 - \theta_m)'\Gamma' E[z_i z_i']\Gamma(\theta_0 - \theta_m)}{\sigma^2_{\eta_m}} \tag{81}$$

$$= \frac{E[(\tilde{\Delta}_{0i} - \tilde{\Delta}_{mi})^2]}{\sigma^2_{\eta_m}}. \tag{82}$$

(iii) - RV To derive the plim for T^{RV}, first consider the object $Q_1 - Q_2$:

$$\begin{aligned}Q_1 - Q_2 &= g_1' W g_1 - g_2' W g_2\\ &= \frac{1}{n}(\Delta(\theta_0 - \theta_1) + \omega_0)' P_z(\Delta(\theta_0 - \theta_1) + \omega_0) - \frac{1}{n}(\Delta(\theta_0 - \theta_2) + \omega_0)' P_z(\Delta(\theta_0 - \theta_2) + \omega_0)\\ &\to_p E[(\tilde{\Delta}_{0i} - \tilde{\Delta}_{1i})^2] - E[(\tilde{\Delta}_{0i} - \tilde{\Delta}_{2i})^2]\end{aligned}$$

So that, assuming $\hat{\sigma}_{RV} \to_p \sigma_{RV}$, we have

$$\frac{T^{RV}}{\sqrt{n}} = \frac{Q_1 - Q_2}{\hat{\sigma}_{RV}} \tag{83}$$

$$\to_p \frac{E[(\tilde{\Delta}_{0i} - \tilde{\Delta}_{1i})^2] - E[(\tilde{\Delta}_{0i} - \tilde{\Delta}_{2i})^2]}{\sigma_{RV}} \tag{84}$$

which is consistent with Theorem 1 (ii) in Rivers and Vuong (2002).

(iv) - Cox For the Cox procedure, notice that the numerator of the test statistic is

$$Q_{-m} - Q_m - (g_{-m} - g_m)' A_{zz}^{-1}(g_{-m} - g_m),$$

for which we obtain

$$
\begin{aligned}
&g_{-m}' A_{zz}^{-1} g_{-m} - g_m' A_{zz}^{-1} g_m - (g_{-m} - g_m)' A_{zz}^{-1} (g_{-m} - g_m) \\
&\rightarrow_p (\theta_0 - \theta_{-m})' A_{z\Delta}' A_{zz}^{-1} A_{z\Delta} (\theta_0 - \theta_{-m}) - (\theta_0 - \theta_m)' A_{z\Delta}' A_{zz}^{-1} A_{z\Delta} (\theta_0 - \theta_m) \\
&\qquad - (\theta_m - \theta_{-m})' A_{z\Delta}' A_{zz}^{-1} A_{z\Delta} (\theta_m - \theta_{-m}) \\
&= E[(\tilde{\Delta}_{0i} - \tilde{\Delta}_{-mi})^2] - E[(\tilde{\Delta}_{0i} - \tilde{\Delta}_{mi})^2] - E[(\tilde{\Delta}_{mi} - \tilde{\Delta}_{-mi})^2] \\
&= 2E[(\tilde{\Delta}_{0i} - \tilde{\Delta}_{mi})(\tilde{\Delta}_{-mi} - \tilde{\Delta}_{mi})]
\end{aligned}
$$

So that, assuming $\hat{\sigma}_{Cox} \rightarrow_p \sigma_{Cox}$, we have

$$
\frac{T^{Cox,m}}{\sqrt{n}} \rightarrow_p \frac{2E[(\tilde{\Delta}_{0i} - \tilde{\Delta}_{mi})(\tilde{\Delta}_{-mi} - \tilde{\Delta}_{mi})]}{\sigma_{Cox}}. \tag{85}
$$

To prove the second part of the proposition, that the tests are consistent, notice that for each test y, there exists an increasing function of n, a_n such that $\frac{T^{y,m}}{a_n} \rightarrow_p \zeta$. The quantity ζ is positive and finite when $\theta_m \neq \theta_0$ and $E(\tilde{\Delta}_{0i}) \neq E(\tilde{\Delta}_{mi})$. Hence, $T^{y,m} \rightarrow_p \infty$, so that each test has power of one against fixed alternatives.

A.4 Proof of Proposition 2

Remarks on Local Alternatives - We start with some preliminary discussion of our definition of local alternatives. We assume in Proposition 2 that:

$$
\hat{A}_{z\Delta} = \frac{z'\Delta}{n} = \frac{\rho}{\sqrt{n}} + o_p(n^{-1/2}), \tag{86}
$$

where ρ is a finite matrix, which intuitively means that the covariation between instruments and markups implied by the two models vanishes as the sample size increases. In the spirit of local power analysis, this is not a restriction to take literally, but rather to use as a tool to approximate the finite sample power of the tests we consider in a challenging environment where the two models are hard to tell apart.

Most standard local power analysis is done by imposing that the parameters that characterize the alternative hypothesis get closer to the truth at a certain rate. In the spirit of the literature on non-nested hypothesis testing and model testing (e.g. Pesaran, 1982; Smith, 1992), we adopt instead a restriction on the joint distribution of the data (z, Δ), imposing that a certain moment vanishes at a rate that allows to control the asymptotic distribution of the test statistics under the alternative hypothesis.

It is worth noticing some of the implications of our assumption (86). First, notice that our definition of local alternatives implies that:

$$
g_m = \frac{z'\omega_m}{n} = \frac{g}{\sqrt{n}} + o_p(n^{-1/2}). \tag{87}
$$

where g is a finite d_z-vector. In fact consider:

$$g_m = \frac{z'(\omega_0 + \Delta(\theta_0 - \theta_m))}{n} \tag{88}$$
$$= \frac{z'\Delta}{n}(\theta_0 - \theta_m) + o_p(1) \tag{89}$$

Then, if (86) holds, it follows immediately that $g_m = \frac{\rho}{\sqrt{n}}(\theta_0 - \theta_m) + o_p(n^{-1/2})$, which implies that (87) holds for $g = \rho(\theta_0 - \theta_m)$. The condition (87) is similar to the definition of local alternatives adopted in Marmer and Otsu (2012).

Notice moreover that we can write the estimator of the first stage parameter $\hat{\Gamma}$ as:

$$\hat{\Gamma} = (\frac{z'z}{n})^{-1}\frac{z'\Delta}{n},$$

so that its population equivalent is

$$\Gamma = \Gamma_n = A_{zz}^{-1}\frac{g}{\sqrt{n}},$$

where the last equality follows from assumption (86). Hence, our assumption, though stronger, implies the condition on Γ routinely used to analyze asymptotics with weak instruments (Stock and Staiger, 1997).

Proof of Proposition 2. We will use in what follows the following facts, which are a consequence of the assumptions of the proposition:

$$\sqrt{n}\hat{A}_{\Delta z} = \rho + o_p(1) \tag{90}$$
$$\hat{A}_{zz} = A_{zz} + o_p(1) \tag{91}$$
$$\frac{z'\omega_0}{\sqrt{n}} \to_d N(0, A_{zz}\sigma^2_{\omega_0}) \tag{92}$$
$$\hat{\sigma}^2_{\omega_0} = \sigma^2_{\omega_0} + o_p(1) \tag{93}$$

The latter three statements follow from the standard assumptions required for the consistency and asymptotic normality of the GMM estimator $\hat{\theta}$, and homoscedasticity of the error terms. To derive the non-centrality parameter, it is also useful to note:

$$\sqrt{n}\hat{A}_{\Delta z}\hat{A}_{zz}^{-1} = \Gamma_n + o_p(1) \tag{94}$$
$$\sqrt{n}\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}z = \tilde{\Delta}_n + o_p(1) \tag{95}$$
$$\sqrt{n}\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\sqrt{n}\hat{A}_{z\Delta} = \sqrt{n}\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\hat{A}_{zz}\hat{A}_{zz}^{-1}\sqrt{n}\hat{A}_{z\Delta} \tag{96}$$
$$= E[\rho' A_{zz}^{-1} z_i z_i' A_{zz}^{-1}\rho] + o_p(1) \tag{97}$$
$$= E[\tilde{\Delta}_{ni}\tilde{\Delta}_{ni}'] + o_p(1) \tag{98}$$

These statements are either definitions, or follow from the assumptions above.

We assume in the proposition that the true parameter is θ_0 and the null hypothesis for AR, EB and Cox is model m such that $\theta_m \neq \theta_0$. We also define the constant:

$$c = (\theta_0 - \theta_m)' E[\tilde{\Delta}_{ni} \tilde{\Delta}'_{ni}](\theta_0 - \theta_m). \tag{99}$$

(i) - EB For EB, notice that - since under our assumptions the variance of the estimator is

$$\hat{V}_\theta = (\hat{A}_{\Delta z} \hat{A}_{zz}^{-1} \hat{A}_{z\Delta})^{-1} \hat{\sigma}^2_{\omega_0} \tag{100}$$

the Wald test statistic for the null $\theta_0 = \theta_m$ can be written as :

$$T^{EB,m} = \frac{\zeta' \zeta}{\hat{\sigma}^2_{\omega_0} / \sigma^2_{\omega_0}} \tag{101}$$

where

$$\zeta = \frac{(\hat{A}_{\Delta z} \hat{A}_{zz}^{-1} \hat{A}_{z\Delta})^{1/2}}{\sigma_{\omega_0}} (\sqrt{n}(\hat{\theta} - \theta_m)) \tag{102}$$

Notice moreover that

$$\begin{aligned} \hat{\theta} &= (\Delta' P_z \Delta)^{-1} \Delta' P_z p \\ &= \theta_0 + (\hat{A}_{\Delta z} \hat{A}_{zz}^{-1} \hat{A}_{z\Delta})^{-1} \hat{A}_{\Delta z} \hat{A}_{zz}^{-1} \frac{z' \omega_0}{n}, \end{aligned}$$

so that

$$\begin{aligned} \zeta &= \frac{(\hat{A}_{\Delta z} \hat{A}_{zz}^{-1} \hat{A}_{z\Delta})^{1/2}}{\sigma_{\omega_0}} (\sqrt{n}(\theta_0 - \theta_m)) + \frac{(\hat{A}_{\Delta z} \hat{A}_{zz}^{-1} \hat{A}_{z\Delta})^{-1/2}}{\sigma_{\omega_0}} \hat{A}_{\Delta z} \hat{A}_{zz}^{-1} \frac{z' \omega_0}{\sqrt{n}} \\ &= \frac{(\sqrt{n} \hat{A}_{\Delta z} \hat{A}_{zz}^{-1} \sqrt{n} \hat{A}_{z\Delta})^{1/2}}{\sigma_{\omega_0}} (\theta_0 - \theta_m) + \frac{(\sqrt{n} \hat{A}_{\Delta z} \hat{A}_{zz}^{-1} \sqrt{n} \hat{A}_{z\Delta})^{-1/2}}{\sigma_{\omega_0}} \sqrt{n} \hat{A}_{\Delta z} \hat{A}_{zz}^{-1} \frac{z' \omega_0}{\sqrt{n}} \end{aligned}$$

from which it follows that, by applying the generalized Slutsky theorem and the facts stated above, that:

$$\zeta \to_d N\left(\frac{(\rho' A_{zz}^{-1} \rho)^{1/2}}{\sigma_{\omega_0}} (\theta_0 - \theta_m), 1 \right).$$

Hence, using the definition of the EB test statistic in Equation (101), we have the following convergence in distribution result:

$$T^{EB} \to_d \chi^2_2 \left(\frac{c}{\sigma^2_{\omega_0}} \right), \tag{103}$$

since

$$(\theta_0 - \theta_m)'\rho' A_{zz}^{-1}\rho(\theta_0 - \theta_m) = c.$$

(ii) - AR For AR, we first consider the estimator $\hat{\pi}_m$

$$\begin{aligned}\hat{\pi}_m &= (z'z)^{-1}z'(p - \Delta\theta_m)\\ &= \hat{A}_{zz}^{-1}\hat{A}_{z'\Delta}(\theta_0 - \theta_m) + \hat{A}_{zz}^{-1}\frac{z'\omega_0}{n},\end{aligned}$$

and, under the assumption of homoskedasticity, $\hat{V}_{\pi_m}$, a consistent estimator of the asymptotic variance V_{π_m}, is:

$$V_{\pi_m} = \sigma_{\omega_0}^2 A_{zz}^{-1}. \tag{104}$$

So, we have

$$\begin{aligned}\sqrt{n}\hat{V}_{\pi_m}^{-1/2}\hat{\pi}_m &= \left(\frac{\hat{A}_{zz}^{-1/2}\sqrt{n}\hat{A}'_{z\Delta}(\theta_0 - \theta_m) + \hat{A}_{zz}^{-1/2}\frac{z'\omega_0}{\sqrt{n}}}{\hat{\sigma}_{\omega_0}}\right) &(105)\\ &\to_d \mathcal{N}\left(\frac{A_{zz}^{-1/2}\rho(\theta_0 - \theta_m)}{\sigma_{\omega_0}}, 1\right), &(106)\end{aligned}$$

and we can derive the asymptotic distribution of the AR test statistic:

$$\begin{aligned}T^{AR,m} &\to_d \chi^2_{d_z}\left(\frac{(\theta_0 - \theta_m)'\rho A_{zz}^{-1}\rho(\theta_0 - \theta_m)}{\sigma_{\omega_0}^2}\right) &(107)\\ &= \chi^2_{d_z}\left(\frac{c}{\sigma_{\omega_0}^2}\right). &(108)\end{aligned}$$

(iii) - RV First consider the object :

$$\begin{aligned}Q_1 - Q_2 &= g_1'Wg_1 - g_2'Wg_2\\ &= \frac{\omega^{I\prime}z}{n}(\frac{z'z}{n})^{-1}\frac{z'\omega^I}{n} - \frac{\omega^{II\prime}z}{n}(\frac{z'z}{n})^{-1}\frac{z'\omega^{II}}{n}\\ &= \frac{1}{n}(\Delta(\theta_0 - \theta_1) + \omega_0)'P_z(\Delta(\theta_0 - \theta_1) + \omega_0) - \frac{1}{n}(\Delta(\theta_0 - \theta_2) + \omega_0)'P_z(\Delta(\theta_0 - \theta_2) + \omega_0)\\ &= -(\theta_1 - \theta_2)'\hat{A}'_{z\Delta}\hat{A}_{zz}^{-1}\hat{A}_{z\Delta}(\theta_1 - \theta_2) - 2(\theta_2 - \theta_1)'\hat{A}'_{z\Delta}\hat{A}_{zz}^{-1}\frac{z'\omega_0}{n}.\end{aligned}$$

Under the assumption of homoskedasticity, the estimator of the asymptotic variance of $\sqrt{n}(Q_1 - Q_2)$ is:

$$\hat{\sigma}_{RV}^2 = 4\hat{\sigma}_{\omega_0}^2(\theta_2 - \theta_1)'\hat{A}'_{z\Delta}\hat{A}_{zz}^{-1}\hat{A}_{z\Delta}(\theta_2 - \theta_1), \tag{109}$$

So that we can write:

$$
\begin{aligned}
T^{RV} &= \sqrt{n}\frac{Q_1 - Q_2}{\hat{\sigma}_{RV}} && (110)\\
&= -\frac{1}{2\hat{\sigma}_{\omega_0}}\Big((\theta_2-\theta_1)'\sqrt{n}\hat{A}'_{z\Delta}\hat{A}_{zz}^{-1}\sqrt{n}\hat{A}_{z\Delta}(\theta_2-\theta_1)\Big)^{1/2} && (111)\\
&\quad - \frac{1}{\hat{\sigma}_{\omega_0}}\Big((\theta_2-\theta_1)'\sqrt{n}\hat{A}'_{z\Delta}\hat{A}_{zz}^{-1}\sqrt{n}\hat{A}_{z\Delta}(\theta_2-\theta_1)\Big)^{-1/2}(\theta_2-\theta_1)'\sqrt{n}\hat{A}'_{z\Delta}\hat{A}_{zz}^{-1}\frac{z'\omega_0}{\sqrt{n}} && (112)
\end{aligned}
$$

For convergence in distribution, after applying the CLT, we have:

$$
\begin{aligned}
T^{RV} &\to_d n\left(\frac{\Big((\theta_2-\theta_1)'E[\tilde{\Delta}_{ni}\tilde{\Delta}'_{ni}](\theta_2-\theta_1)\Big)^{1/2}}{2\sigma_{\omega_0}}, 1\right) && (113)\\
&= n\left(\frac{\Big(E[(\tilde{\Delta}_{2ni}-\tilde{\Delta}_{1ni})^2]\Big)^{1/2}}{2\sigma_{\omega_0}}, 1\right) && (114)\\
&= n\left(\frac{\sqrt{c}}{2\sigma_{\omega_0}}, 1\right). && (115)
\end{aligned}
$$

Therefore

$$
(T^{RV})^2 \to_d \chi_1^2\left(\frac{c}{4\sigma^2_{\omega_0}}\right).
$$

(iv) - Cox To derive the asymptotic distribution of the Cox test statistic, recall that $\theta_0 = \theta_{-m}$. Since the numerator of the test statistic is

$$
\frac{\hat{\sigma}^{Cox,m}}{\sqrt{n}}T^{Cox,m} = Q_{-m} - Q_m - (g_{-m}-g_m)'\hat{A}_{zz}^{-1}(g_{-m}-g_m),
$$

so that the first two terms equal the numerator of the RV test statistic. We conjecture (and verify at the end) that in this environment the asymptotic variance of the Cox test statistic is equal to the one of the RV test statistic, so that we can write

$$
T^{Cox,m} - T^{RV} = -\frac{\sqrt{n}}{\hat{\sigma}^{RV}}(g_{-m}-g_m)'\hat{A}_{zz}^{-1}(g_{-m}-g_m),
$$

and for this term notice

$$
\begin{aligned}
-\frac{\sqrt{n}}{\hat{\sigma}^{RV}}(g_{-m}-g_m)'\hat{A}_{zz}^{-1}(g_{-m}-g_m) &= -\frac{\sqrt{n}}{\hat{\sigma}^{RV}}(\theta_m-\theta_{-m})'\hat{A}_{z\Delta}'\hat{A}_{zz}^{-1}\hat{A}_{z\Delta}(\theta_m-\theta_{-m}) \\
&= -\frac{1}{2\hat{\sigma}_{\omega_0}}\left((\theta_m-\theta_{-m})'\sqrt{n}\hat{A}_{z\Delta}'\hat{A}_{zz}^{-1}\sqrt{n}\hat{A}_{z\Delta}(\theta_m-\theta_{-m})\right)^{1/2} \\
&\to_p -\frac{1}{2\hat{\sigma}_{\omega_0}}\left((\theta_m-\theta_{-m})'E[\tilde{\Delta}_{ni}\tilde{\Delta}_{ni}'](\theta_m-\theta_{-m})\right)^{1/2} \\
&= -\frac{c^{1/2}}{2\hat{\sigma}_{\omega_0}}
\end{aligned}
$$

So that, since this term converges in probability to a constant, the asymptotic variance for the Cox statistic that we conjecture is correct. In fact:

$$
T^{Cox,m} \to_d N\left(\frac{c^{1/2}}{\hat{\sigma}_{\omega_0}}, 1\right) \tag{116}
$$

Following the results above, we obtain

$$
(T^{cox})^2 \to_d \chi_1^2\left(\frac{c}{\sigma_{\omega_0}^2}\right). \tag{117}
$$

$\square$

A.5 Proof of Proposition 3

Useful objects for the proof:

$$
\begin{aligned}
\hat{A}_{\Delta\Delta} &= \frac{\Delta'\Delta}{n} \\
\hat{A}_{\Delta p} &= \frac{\Delta' p}{n}
\end{aligned}
$$

and the facts:

$$
\begin{aligned}
\hat{A}_{\Delta\Delta} &= A_{\Delta\Delta} + o_p(1) \\
\hat{A}_{\Delta p} &= A_{\Delta p} + o_p(1).
\end{aligned}
$$

Notice moreover that:

$$
E[\hat{\theta}_{OLS}] = A_{\Delta\Delta}^{-1}A_{\Delta p} = \theta_{OLS} \neq \theta_0.
$$

where the inequality comes from our assumption on the endogeneity of the markup term. We first prove a useful lemma:

Lemma 3. *Define the RSS measure of fit as*

$$Q^{RSS}(\theta) = \frac{1}{n}\sum \hat{\omega}_{jt}(\theta)^2. \tag{118}$$

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

$$Q^{OLS}(\theta) = h(\theta)'Wh(\theta) \tag{119}$$

for

$$h(\theta) = \frac{1}{n}\Delta'\hat{\omega}(\theta), \tag{120}$$

and W *equals to* $(\Delta'\Delta)^{-1}$*. Then, the two functions* $Q^{RSS}(\theta)$ *and* $Q^{OLS}(\theta)$ *are equal up to a constant, and have the same minimizers. It follows that:*

$$T^{RSS} = \sqrt{n}\frac{Q_1^{OLS} - Q_2^{OLS}}{\hat{\sigma}_{RSS}}$$

Proof of Lemma 2. We can rewrite (119) in matrix notation as:

$$\begin{aligned}
n \cdot Q^{OLS}(\theta) &= \hat{\omega}(\theta)'\Delta(\Delta'\Delta)^{-1}\Delta'\hat{\omega}(\theta) && (121)\\
&= p'\Delta(\Delta'\Delta)^{-1}\Delta'p + (\theta'\Delta'\Delta\theta) - 2p'\Delta\theta && (122)\\
&= p'p - p'p + p'\Delta(\Delta'\Delta)^{-1}\Delta'p + (\theta'\Delta'\Delta\theta) - 2p'\Delta\theta && (123)\\
&= n \cdot Q^{RSS}(\theta) - (p'p - p'\Delta(\Delta'\Delta)^{-1}\Delta'p) && (124)\\
&= n \cdot Q^{RSS}(\theta) + c_n. && (125)
\end{aligned}$$

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

$$\begin{aligned}
Q^{RSS}(\theta_1) - Q^{RSS}(\theta_2) &= Q_1^{RSS} - Q_2^{RSS} && (126)\\
&= Q_1^{OLS} - Q_2^{OLS}. && (127)
\end{aligned}$$

□

Proof of Proposition 3:

By Lemma 1, the RSS test statistic is:

$$T^{RSS} = \sqrt{n}\frac{Q_1^{OLS} - Q_2^{OLS}}{\hat{\sigma}_{RSS}} \tag{128}$$

where $\hat{\sigma}_{I,II}^{RSS} \to_p \sqrt{\sigma_{I,II}^{2,RSS}}$ and $\sigma_{I,II}^{2RSS}$ is the asymptotic variance of $\sqrt{n}(Q_1^{OLS} - Q_2^{OLS})$. To derive the plim of the test statistic, first consider the object:

$$
\begin{aligned}
Q_1^{OLS} - Q_2^{OLS} &= g_1' W g_1 - n g_2' W g_2 \\
&= -(\theta_1 - \theta_2)' \hat{A}_{\Delta\Delta}(\theta_1 - \theta_2) - 2(\theta_1 - \theta_2)' \frac{\Delta' \omega_0}{n} \\
&\to_p -(\theta_1 - \theta_2)' A_{\Delta\Delta}(\theta_1 - \theta_2) - 2(\theta_1 - \theta_2)' E[\Delta' \omega_0] \\
&= -E[(\Delta_{1i} - \Delta_{2i})^2] - 2(\theta_1 - \theta_2)' E[\Delta_i \Delta_i'](\theta_{OLS} - \theta_0) \\
&= -E[(\Delta_{1i} - \Delta_{2i})^2] - 2E[(\Delta_{1i} - \Delta_{2i})'(\Delta_{OLS,i} - \Delta_{1i})] \\
&= E[(\Delta_{OLS,i} - \Delta_{1i}))^2 - (\Delta_{OLS,i} - \Delta_{2i}))^2],
\end{aligned}
$$

so that we have

$$
\frac{T^{RSS}}{\sqrt{n}} \to_p \frac{E[(\Delta_{OLS,i} - \Delta_{1i}))^2 - (\Delta_{OLS,i} - \Delta_{2i}))^2]}{\sigma^{RSS}_{I,II}}. \tag{129}
$$

A.6 Proof of Proposition 4

It is useful to note the following identities for each model m:

$$
\omega_m = p - \Delta\theta_m \tag{130}
$$
$$
\omega_m = \Delta^*\theta_0 - \Delta\theta_m + \omega_0 \tag{131}
$$
$$
\omega_m = \Delta^*_0 - \Delta_m + \omega_0. \tag{132}
$$

Since in this environment true markups Δ^* differ from observed markups Δ, we need the following definitions:

$$
A_{\Delta z} A_{zz}^{-1} A_{z\Delta^*} = E[\tilde{\Delta}_i \tilde{\Delta}_i^{*\prime}], \tag{133}
$$

and

$$
\lambda = E[\tilde{\Delta}_i \tilde{\Delta}_i']^{-1} E[\tilde{\Delta}_i \tilde{\Delta}_i^{*\prime}]. \tag{134}
$$

(i) - EB For EB, we first consider the object $\hat{\theta}$

$$
\begin{aligned}
\hat{\theta} &= (\Delta' P_z \Delta)^{-1} \Delta' P_z p \\
&= (\hat{A}_{\Delta z} \hat{A}_{zz}^{-1} \hat{A}_{z\Delta})^{-1} \hat{A}_{\Delta z} \hat{A}_{zz}^{-1} \hat{A}_{z\Delta^*} \theta_0 + (\hat{A}_{\Delta z} \hat{A}_{zz}^{-1} \hat{A}_{z\Delta})^{-1} \hat{A}_{\Delta z} \hat{A}_{zz}^{-1} \frac{z' \omega_0}{n} \\
&\to_p (A_{\Delta z} A_{zz}^{-1} A_{z\Delta})^{-1} A_{\Delta z} A_{zz}^{-1} A_{z\Delta^*} \theta_0 + (A_{\Delta z} A_{zz}^{-1} A_{z\Delta})^{-1} A_{\Delta z} A_{zz}^{-1} E[z_i \omega_{0i}'] \\
&= \lambda \theta_0
\end{aligned}
$$

So we have, approximating the asymptotic variance with the expression (66)[30]:

$$\begin{aligned}
n(\hat{\theta}-\theta_m)'\hat{V}_\theta^{-1}(\hat{\theta}-\theta_m) &\to_p \lim_{n\to\infty} n((\lambda\theta_0-\theta_m)'V_\theta^{-1}(\lambda\theta_0-\theta_m)) && (135)\\
&\approx \lim_{n\to\infty} n\frac{(\lambda\theta_0-\theta_m)'E[\tilde{\Delta}_i\tilde{\Delta}_i'](\lambda\theta_0-\theta_m)}{\sigma^2_{\omega_0}} && (136)\\
&= \lim_{n\to\infty} n\frac{E[(\theta_0'\lambda\tilde{\Delta}_i-\tilde{\Delta}_{mi})^2]}{\sigma^2_{\omega_0}} && (137)\\
&= \lim_{n\to\infty} n\frac{E\left[\left((\tilde{\tilde{\Delta}}^*_{0i}-\tilde{\Delta}_{0i})+(\tilde{\Delta}_{0i}-\tilde{\Delta}_{mi})\right)^2\right]}{\sigma^2_{\omega_0}}. && (138)
\end{aligned}$$

(ii) - AR To derive the plim of the AR test statistic, we first consider the object $\hat{\pi}_m$

$$\begin{aligned}
\hat{\pi}_m &= (z'z)^{-1}z'(\Delta^*\theta_0+\omega_0-\Delta\theta_m)\\
&= \hat{A}_{zz}^{-1}(\hat{A}_{z\Delta^*}\theta_0-\hat{A}_{z\Delta}\theta_m)+\hat{A}_{zz}^{-1}\frac{z'\omega_0}{n}\\
&\to_p A_{zz}^{-1}(A_{z\Delta^*}\theta_0-A_{z\Delta}\theta_m)+A_{zz}^{-1}E[z_i\omega_{0i}']\\
&= \Gamma^*\theta_0-\Gamma\theta_m
\end{aligned}$$

So we have, under the approximation of the asymptotic variance given in (66):

$$\begin{aligned}
T^{AR,m} &\to_p \lim_{n\to\infty} n(\Gamma^*\theta_0-\Gamma\theta_m)'V_{\pi_m}^{-1}(\Gamma^*\theta_0-\Gamma\theta_m)\\
&\approx \lim_{n\to\infty} n\frac{(\Gamma^*\theta_0-\Gamma\theta_m)'A_{zz}(\Gamma^*\theta_0-\Gamma\theta_m)}{\sigma^2_{\omega_0}}\\
&= \lim_{n\to\infty} n\frac{E[(\tilde{\Delta}^*_{0i}-\tilde{\Delta}_{mi})^2]}{\sigma^2_{\omega_0}}\\
&= \lim_{n\to\infty} n\frac{E\left[\left((\tilde{\Delta}^*_{0i}-\tilde{\Delta}_{0i})+(\tilde{\Delta}_{0i}-\tilde{\Delta}_{mi})\right)^2\right]}{\sigma^2_{\omega_0}}
\end{aligned}$$

[30]Since the proposition concerns the plim of the test statistic, adopting this approximation does not alter the result.

(iii) - RV To derive the plim of the RV test statistic we start from the object:

$$\begin{aligned}
Q_1 - Q_2 &= g_1'Wg_1 - g_2'Wg_2 \\
&= \frac{1}{n}(\Delta^*\theta_0 - \Delta\theta_1 + \omega_0)'P_z(\Delta^*\theta_0 - \Delta\theta_1 + \omega_0) - \frac{1}{n}(\Delta^*\theta_0 - \Delta\theta_2 + \omega_0)'P_z(\Delta^*\theta_0 - \Delta\theta_2 + \omega_0) \\
&= (\theta_0\hat{A}_{\Delta^*z} - \theta_1\hat{A}_{\Delta z})'\hat{A}_{zz}^{-1}(\hat{A}_{\Delta^*z}\theta_0 - \hat{A}_{\Delta z}\theta_1) - (\theta_0\hat{A}_{\Delta^*z} - \theta_2\hat{A}_{\Delta z})'\hat{A}_{zz}^{-1}(\hat{A}_{\Delta^*z}\theta_0 - \hat{A}_{\Delta z}\theta_2) \\
&\quad + 2(\theta_2 - \theta_1)\hat{A}_{\Delta z}A_{zz}^{-1}\frac{\omega_0'z}{n} \\
&\to_p (\theta_0 A_{\Delta^*z} - \theta_1 A_{\Delta z})'A_{zz}^{-1}A_{zz}A_{zz}^{-1}(A_{\Delta^*z}\theta_0 - A_{\Delta z}\theta_1) \\
&\quad - (\theta_0 A_{\Delta^*z} - \theta_2 A_{\Delta z})'A_{zz}^{-1}A_{zz}A_{zz}^{-1}(A_{\Delta^*z}\theta_0 - A_{\Delta z}\theta_2) \\
&= E[(\theta_0\tilde{\Delta}_i^* - \theta_1\tilde{\Delta}_i)^2] - E[(\theta_0\tilde{\Delta}_i^* - \theta_2\tilde{\Delta}_i)^2]
\end{aligned}$$

So we have, under the approximation of the asymptotic variance given in (66):

$$T^{RV} \to_p \lim_{n\to\infty}\sqrt{n}\frac{E[(\theta_0\tilde{\Delta}_i^* - \theta_1\tilde{\Delta}_i)^2] - E[(\theta_0\tilde{\Delta}_i^* - \theta_2\tilde{\Delta}_i)^2]}{\sigma_{RV}} \tag{139}$$

$$\approx \lim_{n\to\infty}\sqrt{n}\frac{E[(\theta_0\tilde{\Delta}_i^* - \theta_1\tilde{\Delta}_i)^2] - E[(\theta_0\tilde{\Delta}_i^* - \theta_2\tilde{\Delta}_i)^2]}{2\sigma_{\omega_0}\sqrt{E[(\tilde{\Delta}_{1i} - \tilde{\Delta}_{2i})^2]]}} \tag{140}$$

$$= \lim_{n\to\infty}\sqrt{n}\frac{E\left[\left((\tilde{\Delta}_{0i}^* - \tilde{\Delta}_{1i}^*) + (\tilde{\Delta}_{1i}^* - \tilde{\Delta}_{1i})\right)^2 - \left((\tilde{\Delta}_{0i}^* - \tilde{\Delta}_{2i}^*) + (\tilde{\Delta}_{2i}^* - \tilde{\Delta}_{2i})\right)^2\right]}{2\sigma_{\omega_0}\sqrt{E[(\tilde{\Delta}_{1i} - \tilde{\Delta}_{2i})^2]}}. \tag{141}$$

(v) - Cox For Cox, the numerator of the test statistic when using the formulation in Smith (1992) is:

$$\begin{aligned}
&g_{-m}'A_{zz}^{-1}g_{-m} - g_m'A_{zz}^{-1}g_m - (g_{-m} - g_m)'A_{zz}^{-1}(g_{-m} - g_m) \\
&= Q_{-m} - Q_m - (\theta_m'\hat{A}_{z\Delta*}' - \theta_{-m}'\hat{A}_{z\Delta}')\hat{A}_{z'z}^{-1}(\hat{A}_{z\Delta*}\theta_m - \hat{A}_{z\Delta}\theta_{-m})) \\
&\to_p E[(\tilde{\Delta}_{0i}^* - \tilde{\Delta}_{-mi})^2] - E[(\tilde{\Delta}_{0i}^* - \tilde{\Delta}_{mi})^2] - E[(\tilde{\Delta}_{mi}^* - \tilde{\Delta}_{-mi})^2]
\end{aligned}$$

So we have, under the approximation of the asymptotic variance given in (66):

$$T^{cox,m} \to_p 2\lim_{n\to\infty}\sqrt{n}\frac{E[(\tilde{\Delta}_{0i}^* - \tilde{\Delta}_{-mi})^2] - E[(\tilde{\Delta}_{0i}^* - \tilde{\Delta}_{mi})^2] - E[(\tilde{\Delta}_{mi}^* - \tilde{\Delta}_{-mi})^2]}{\sigma_{cox}} \tag{142}$$

$$\approx 2\lim_{n\to\infty}\sqrt{n}\frac{E[(\tilde{\Delta}_{0i}^* - \tilde{\Delta}_{-mi})^2] - E[(\tilde{\Delta}_{0i}^* - \tilde{\Delta}_{mi})^2] - E[(\tilde{\Delta}_{mi}^* - \tilde{\Delta}_{-mi})^2]}{2\sigma_{\omega_0}\sqrt{E[(\tilde{\Delta}_{1i} - \tilde{\Delta}_{2i})^2]}}. \tag{143}$$

A.7 Proof of Proposition 5

On top of the assumption of Proposition 2 that $\sqrt{n}\hat{A}_{z\Delta} = \rho + o_p(1)$, we also need a further assumption to define local alternatives, this time involving the true markups Δ^*:

$$\sqrt{n}(\hat{A}_{z\Delta^*} - \hat{A}_{z\Delta}) = \rho^* + o_p(1), \tag{144}$$

so that:

$$\sqrt{n}(\hat{A}_{z\Delta^*}\theta_0 - \hat{A}_{z\Delta}\theta_m) = A_{zz}^{-1}\rho^*\theta_0 + A_{zz}^{-1}\rho(\theta_0 - \theta_m) + o_p(1). \tag{145}$$

We formally state the assumption as:

Assumption 10. *We assume that the models we consider satisfy the following local alternatives property:*

$$\sqrt{n}(\hat{A}_{z\Delta^*} - \hat{A}_{z\Delta}) = \rho^* + o_p(1), \tag{146}$$

where ρ^* is a finite matrix.

It is also useful to note the following:

$$\sqrt{n}(\hat{A}_{z\Delta^*}\theta_0 - \hat{A}_{z\Delta}\theta_m)\hat{A}_{zz}^{-1}\sqrt{n}(\hat{A}_{z\Delta^*}\theta_0 - \hat{A}_{z\Delta}\theta_m) = E\left[\left(\theta_0\tilde{\Delta}^*_{ni} - \theta_m\tilde{\Delta}_{ni}\right)\left(\tilde{\Delta}^*_{ni}\theta_0 - \tilde{\Delta}_{ni}\theta_m\right)\right] + o_p(1) \tag{147}$$

$$= E\left[\left(\tilde{\Delta}^*_{n,0i} - \tilde{\Delta}_{n,mi}\right)^2\right] + o_p(1) \tag{148}$$

(i) - EB To derive the distribution of the test statistic, we first consider:

$$\begin{aligned}
\sqrt{n}(\Delta' P_z\Delta)^{1/2}(\hat{\theta} - \theta_0) &= \sqrt{n}(\Delta' P_z\Delta)^{-1/2}\Delta' P_z p - \sqrt{n}(\Delta' P_z\Delta)^{1/2}\theta_0 \\
&= \sqrt{n}(\Delta' P_z\Delta)^{-1/2}\Delta' P_z(\Delta^*\theta_0 + \omega_0) - \sqrt{n}(\Delta' P_z\Delta)^{1/2}\theta_0 \\
&= \sqrt{n}(\Delta' P_z\Delta)^{-1/2}\Delta' P_z(\Delta^* - \Delta)\theta_0 + \sqrt{n}(\Delta' P_z\Delta)^{-1/2}\Delta' P_z\omega_0 \\
&= (\sqrt{n}\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\sqrt{n}\hat{A}_{z\Delta})^{-1/2}\sqrt{n}\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\sqrt{n}(\hat{A}_{z\Delta^*} - \hat{A}_{z\Delta})\theta_0 \\
&\quad + (\sqrt{n}\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\sqrt{n}\hat{A}_{z\Delta})^{-1/2}\sqrt{n}\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\frac{z'\omega_0}{\sqrt{n}} \\
&\to_d N\left((\rho A_{zz}^{-1}\rho)^{-1/2}\rho A_{zz}^{-1}\rho^*\theta_0, \sigma^2_{\omega_0}\right)
\end{aligned}$$

so that

$$
\begin{aligned}
\sqrt{n}\frac{(\Delta' P_z \Delta)^{1/2}}{\hat{\sigma}_{\omega_0}}(\hat{\theta}-\theta_m) &= \sqrt{n}\frac{(\Delta' P_z \Delta)^{1/2}}{\hat{\sigma}_{\omega_0}}(\hat{\theta}-\theta_0) + \sqrt{n}\frac{(\Delta' P_z \Delta)^{1/2}}{\hat{\sigma}_{\omega_0}}(\theta_0-\theta_m) \\
&= \sqrt{n}\frac{(\sqrt{n}\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\sqrt{n}\hat{A}_{z\Delta})^{1/2}}{\hat{\sigma}_{\omega_0}}(\hat{\theta}-\theta_0) + \frac{(\sqrt{n}\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\sqrt{n}\hat{A}_{z\Delta})^{1/2}}{\hat{\sigma}_{\omega_0}}(\theta_0-\theta_m) \\
&\to_d N\left(\frac{(\rho A_{zz}^{-1}\rho)^{-1/2}\rho A_{zz}^{-1}\rho^*\theta_0 + (\rho A_{zz}^{-1}\rho)^{1/2}(\theta_0-\theta_m)}{\hat{\sigma}_{\omega_0}}, 1\right) \\
&\to_d N\left(\frac{(\rho A_{zz}^{-1}\rho)^{1/2}((1+\lambda)\theta_0-\theta_m)}{\hat{\sigma}_{\omega_0}}, 1\right)
\end{aligned}
$$

where $\lambda = (\rho A_{zz}^{-1}\rho)^{-1}\rho A_{zz}^{-1}\rho^*$. Thus,

$$
\begin{aligned}
T^{EB} &\to_d \chi_2^2\left(\frac{((1+\lambda)\theta_0-\theta_m)'(\rho A_{zz}^{-1}\rho)((1+\lambda)\theta_0-\theta_m)}{\sigma_{\omega_0}^2}\right) && (149) \\
&= \chi_2^2\left(\frac{((1+\lambda)\theta_0-\theta_m)'E[\tilde{\Delta}_{ni}\tilde{\Delta}_{ni}'](((1+\lambda)\theta_0-\theta_m)}{\sigma_{\omega_0}^2}\right) && (150)
\end{aligned}
$$

(ii) - AR To derive the distribution of the test statistic, we first consider:

$$
\begin{aligned}
\sqrt{n}\hat{\pi}_m &= \sqrt{n}(z'z)^{-1}z'(p-\Delta\theta_m) \\
&= \sqrt{n}(z'z)^{-1}z'(\Delta^*\theta_0+\omega_0-\Delta\theta_m) \\
&= \hat{A}_{zz}^{-1}\sqrt{n}(\hat{A}_{z'\Delta^*}\theta_0-\hat{A}_{z'\Delta}\theta_m) + \hat{A}_{zz}^{-1}\frac{z'\omega_0}{\sqrt{n}} \\
&\to_d \mathcal{n}\left(A_{zz}^{-1}\rho^*\theta_0 + A_{zz}^{-1}\rho(\theta_0-\theta_m), A_{zz}^{-1}\sigma_{\omega_0}^2\right)
\end{aligned}
$$

Thus,

$$
\begin{aligned}
T^{AR} &\to_d \chi_{d_z}^2\left(\frac{(\rho^*\theta_0+\rho(\theta_0-\theta_m))'A_{zz}^{-1}(\rho^*\theta_0+\rho(\theta_0-\theta_m))}{\sigma_{\omega_0}^2}\right) && (151) \\
&= \chi_{d_z}^2\left(\frac{E[(\tilde{\Delta}_{n,0i}^*-\tilde{\Delta}_{n,mi})^2]}{\sigma_{\omega_0}^2}\right) && (152)
\end{aligned}
$$

(iii) - RV Define

$$
\hat{B} = \left((\theta_2-\theta_1)'\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\hat{A}_{z\Delta}(\theta_2-\theta_1)\right)^{-1/2}
$$

To derive the distribution of the test statistic, we first consider:

$$
\begin{aligned}
\sqrt{n}\hat{B}(Q_1 - Q_2) &= \sqrt{n}\hat{B}\frac{1}{n}(\Delta^*\theta_0 - \Delta\theta_1 + \omega_0)'P_z(\Delta^*\theta_0 - \Delta\theta_1 + \sqrt{n}\hat{B}\omega_0) \\
&\qquad - \sqrt{n}\hat{B}\frac{1}{n}(\Delta^*\theta_0 - \sqrt{n}\hat{B}\Delta\theta_2 + \omega_0)'P_z(\Delta^*\theta_0 - \Delta\theta_2 + \omega_0) \\
&= \Bigg((\theta_2 - \theta_1)'\sqrt{n}\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\sqrt{n}\hat{A}_{z\Delta}(\theta_2 - \theta_1)\Bigg)^{-1/2} \times \\
&\qquad \sqrt{n}(\theta_0\hat{A}_{\Delta^* z} - \theta_1\hat{A}_{\Delta z})'\hat{A}_{zz}^{-1}\sqrt{n}(\hat{A}_{\Delta^* z}\theta_0 - \hat{A}_{\Delta z}\theta_1) \\
&\quad - \Bigg((\theta_2 - \theta_1)'\sqrt{n}\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\sqrt{n}\hat{A}_{z\Delta}(\theta_2 - \theta_1)\Bigg)^{-1/2} \times \\
&\qquad \sqrt{n}(\theta_0\hat{A}_{\Delta^* z} - \theta_2\hat{A}_{\Delta z})'\hat{A}_{zz}^{-1}\sqrt{n}(\hat{A}_{\Delta^* z}\theta_0 - \hat{A}_{\Delta z}\theta_2) \\
&\quad + 2\Bigg((\theta_2 - \theta_1)'\sqrt{n}\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\sqrt{n}\hat{A}_{z\Delta}(\theta_2 - \theta_1)\Bigg)^{-1/2} \times \\
&\qquad (\theta_2 - \theta_1)\sqrt{n}\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\frac{\omega_0' z}{\sqrt{n}} \\
&\to_d \mathcal{N}\Bigg(\Bigg((\theta_2 - \theta_1)'\rho A_{zz}^{-1}\rho(\theta_2 - \theta_1)\Bigg)^{-1/2}\mu^{I,II}, 4\sigma_{\omega_0}^2\Bigg)
\end{aligned}
$$

where

$$
\mu^{I,II} = \Big(\rho^*\theta_0 + \rho(\theta_0 - \theta_1)\Big)' A_{zz}^{-1}\Big(\rho^*\theta_0 + \rho(\theta_0 - \theta_1)\Big) - \Big(\rho^*\theta_0 + \rho(\theta_0 - \theta_2)\Big)' A_{zz}^{-1}\Big(\rho^*\theta_0 + \rho(\theta_0 - \theta_2)\Big)
$$

So that, if we square the RV test statistic, we obtain:

$$
(T^{RV})^2 \to_d \chi_1^2\Bigg(\frac{1}{4\sigma_{\omega_0}^2}\Bigg(\frac{E[(\tilde{\Delta}_{n,0i}^* - \tilde{\Delta}_{n,1i})^2]}{E[(\tilde{\Delta}_{n,1i} - \tilde{\Delta}_{n,2i})^2]^{1/2}} - \frac{E[(\tilde{\Delta}_{n,0i}^* - \tilde{\Delta}_{n,2i})^2]}{E[(\tilde{\Delta}_{n,1i} - \tilde{\Delta}_{n,2i})^2]^{1/2}}\Bigg)^2\Bigg). \tag{153}
$$

(iv) - Cox To derive the distribution of the test statistic consider:

$$
\begin{aligned}
\sqrt{n}\hat{B}T^{Cox,m} &= \sqrt{n}\hat{B}\Bigg(Q_{-m} - Q_m - \frac{1}{n}(g_m - g_{-m})'W(g_m - g_{-m})\Bigg) \\
&= \sqrt{n}\hat{B}(Q_{-m} - Q_m) - \sqrt{n}\hat{B}(\theta_m - \theta_{-m})'\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\hat{A}_{z\Delta}(\theta_m - \theta_{-m}) \\
&= \sqrt{n}\hat{B}(Q_{-m} - Q_m) - \Bigg((\theta_2 - \theta_1)'\sqrt{n}\hat{A}_{\Delta z}A_{zz}^{-1}\sqrt{n}\hat{A}_{z\Delta}(\theta_2 - \theta_1)\Bigg)^{-1/2} \times \\
&\qquad (\theta_m - \theta_{-m})'\sqrt{n}\hat{A}_{\Delta z}\hat{A}_{zz}^{-1}\sqrt{n}\hat{A}_{z\Delta}(\theta_m - \theta_{-m}) \\
&\to_d \mathcal{N}(\Bigg((\theta_m - \theta_{-m})'\rho A_{zz}^{-1}\rho(\theta_m - \theta_{-m})\Bigg)^{-1/2}\mu^{cox,m}, 4\sigma_{\omega_0}^2)
\end{aligned}
$$

where

$$\mu^{cox,m} = \Big(\rho^*\theta_0 + \rho(\theta_0 - \theta_{-m})\Big)' A_{zz}^{-1}\Big(\rho^*\theta_0 + \rho(\theta_0 - \theta_{-m})\Big)$$
$$- \Big(\rho^*\theta_0 + \rho(\theta_0 - \theta_m)\Big)' A_{zz}^{-1}\Big(\rho^*\theta_0 + \rho(\theta_0 - \theta_m)\Big)$$
$$- (\theta_{-m} - \theta_m)'\rho' A_{zz}^{-1}\rho(\theta_{-m} - \theta_m)$$

So that, if we square the Cox test statistic, we obtain:

$$(T^{cox})^2 \to_d \chi_1^2\Bigg(\bigg(\frac{E[(\tilde{\Delta}^*_{n,0i} - \tilde{\Delta}_{n,1i})^2] - E[(\tilde{\Delta}^*_{n,0i} - \tilde{\Delta}_{n,2i})^2] - E[(\tilde{\Delta}_{n,1i} - \tilde{\Delta}_{n,2i})^2]}{2\sigma_{\omega_0} E[(\tilde{\Delta}_{n,1i} - \tilde{\Delta}_{n,2i})^2]^{1/2}}\bigg)^2\Bigg) \tag{154}$$

$$= \chi_1^2\bigg(\frac{c_1^{*2}/c_{1,2} - c_2^{*2}/c_{1,2} - c_{1,2}}{4\sigma_{\omega_0}^2}\bigg) \tag{155}$$

B Two-Step Adjustment

B.1 RV Test

To illustrate the two step adjustment, we consider the unorthogonalized variables without loss. We begin by defining some notation: λ are demand parameters, τ_m are supply parameters under model m as defined above. Let ψ_m be the stacked vector of parameters (λ, τ_m). Demand moments are noted as $h(\tilde{\mathbf{y}}, \lambda)$ and supply moments are defined as $g_m(\tilde{\mathbf{y}}, \psi_m)$ where now we explicitly denote how the supply moments depend on the demand side parameters. Both demand and supply moments are function of the data $\tilde{\mathbf{y}}$, which we suppress below. Further define the population analog of Q_m as $\mathcal{Q}_m$.

We form the first order Taylor expansion of the GMM objective function for model m. Specifically, for each model m, we expand the function around the plim of $\hat{\psi}_m$, which we denote $\tilde{\psi}_m$:

$$Q_m(\hat{\psi}_m) = Q_m(\tilde{\psi}_m) + \nabla_\lambda Q_m(\tilde{\psi}_m)(\hat{\lambda} - \tilde{\lambda}) + \nabla_\tau Q_m(\tilde{\psi}_m)(\hat{\tau}_m - \tilde{\tau}_m) \tag{156}$$

For each of the gradient terms in (156), we can write

$$\nabla_\lambda Q_m(\tilde{\psi}_m)(\hat{\lambda} - \lambda_0) = \nabla_\lambda \mathcal{Q}_m(\tilde{\psi}_m)(\hat{\lambda} - \lambda_0) + o_p(n^{-1/2}) \tag{157}$$

$$\nabla_\tau Q_m(\tilde{\psi}_m)(\hat{\tau}_m - \tilde{\tau}_m) = \nabla_\tau \mathcal{Q}_m(\tilde{\psi}_m)(\hat{\tau}_m - \tilde{\tau}_m) + o_p(n^{-1/2}) \tag{158}$$

Then

$$Q_m(\hat{\psi}_m) = Q_m(\tilde{\psi}_m) + \nabla_\lambda \mathcal{Q}_m(\tilde{\psi}_m)(\hat{\lambda} - \lambda_0) + \nabla_\tau \mathcal{Q}_m(\tilde{\psi}_m)(\hat{\tau}_m - \tilde{\tau}_m) + o_p(n^{-1/2}) \tag{159}$$

Note that, for each model m, $\nabla_\tau Q_m(\hat{\psi}_m) = 0$. Then, $plim(\nabla_\tau Q_m(\hat{\psi}_m)) = \nabla_\tau \mathcal{Q}_m(\tilde{\psi}_m) = 0$, and we can write:

$$\sqrt{n} Q_m(\hat{\psi}_m) = \sqrt{n} Q_m(\tilde{\psi}_m) + \sqrt{n} \nabla_\lambda \mathcal{Q}_m(\tilde{\psi}_m)(\hat{\lambda} - \lambda_0) + o_p(1) \tag{160}$$

With this, we can write

$$\sqrt{n}[Q_1(\hat{\psi}_1) - Q_2(\hat{\psi}_2)] = \sqrt{n}[Q_1(\tilde{\psi}_1) - Q_2(\tilde{\psi}_2)] \tag{161}$$

$$+ \sqrt{n}[\nabla_\lambda \mathcal{Q}_1(\tilde{\psi}_1) - \nabla_\lambda \mathcal{Q}_2(\tilde{\psi}_2)](\hat{\lambda} - \lambda_0) + o_p(1) \tag{162}$$

Since, under H_0, $\mathcal{Q}_1(\tilde{\psi}_1) = \mathcal{Q}_2(\tilde{\psi}_2) \neq 0$ we have

$$\sqrt{n}[Q_1(\hat{\psi}_1) - Q_2(\hat{\psi}_2)] = \sqrt{n}[Q_1(\tilde{\psi}_1) - \mathcal{Q}_1(\tilde{\psi}_1) + \nabla_\lambda \mathcal{Q}_1(\tilde{\psi}_1)(\hat{\lambda} - \lambda_0)] \tag{163}$$

$$- \sqrt{n}[Q_2(\tilde{\psi}_2) - \mathcal{Q}_2(\tilde{\psi}_2) - \nabla_\lambda \mathcal{Q}_2(\tilde{\psi}_2)(\hat{\lambda} - \lambda_0)] + o_p(1) \tag{164}$$

Exploiting the quadratic form structure of Q_m and $\mathcal{Q}_m$, and assuming the weight matrix is the identity matrix, we have:

$$\begin{aligned}\sqrt{n}[Q_1(\hat{\psi}_1) - Q_2(\hat{\psi}_2)] = {} & 2[E[g_1(\tilde{\psi}_1)]'\sqrt{n}(g(\tilde{\psi}_1) - E[g_1(\tilde{\psi}_1)]] \\ & + \nabla_\lambda \mathcal{Q}_1(\tilde{\psi}_1)(\hat{\lambda} - \lambda_0) \\ & - 2[E[g_1(\tilde{\psi}_2)]'\sqrt{n}(g(\tilde{\psi}_2) - E[g_2(\tilde{\psi}_2)]] \\ & - \nabla_\lambda \mathcal{Q}_2(\tilde{\psi}_2)(\hat{\lambda} - \lambda_0) + o_p(1)\end{aligned} \tag{165}$$

Note that since the population GMM objective is a quadratic form, we can write:

$$\nabla_\lambda \mathcal{Q}_m(\tilde{\psi}_m) = 2E[\nabla_\lambda g_m(\tilde{\psi}_m)]'E[g_m(\tilde{\psi}_m)] \tag{166}$$

Plugging (166) into (165) we obtain:

$$\begin{aligned}\sqrt{n}[Q_1(\hat{\psi}_1) - Q_2(\hat{\psi}_2)] = {} & 2\left[E[g_1(\tilde{\psi}_1)]'\sqrt{n}\left(g(\tilde{\psi}_1) - E[g_1(\tilde{\psi}_1)] + E[\nabla_\lambda g_1(\tilde{\psi}_1)](\hat{\lambda} - \lambda_0)\right)\right] \\ & - 2\left[E[g_2(\tilde{\psi}_2)]'\sqrt{n}\left(g(\tilde{\psi}_2) - E[g_2(\tilde{\psi}_2)] + E[\nabla_\lambda g_2(\tilde{\psi}_2)](\hat{\lambda} - \lambda_0)\right)\right] \\ & + o_p(1)\end{aligned} \tag{167}$$

It follows that a consistent estimator of the limiting variance of $\sqrt{n}[Q_1(\hat{\psi}_1) - Q_2(\hat{\psi}_2)]$ is:

$$\hat{\sigma}^2 = 4g_1(\hat{\psi}_1)'\Sigma_1 g_1(\hat{\psi}_1) + 4g_2(\hat{\psi}_2)'\Sigma_2 g_2(\hat{\psi}_2) - 8g_1(\hat{\psi}_1)'\Sigma_{1,2} g_2(\hat{\psi}_2) \tag{168}$$

where

$$\Sigma_m = \sum_{i=1}^{n} \left(\mathbf{z}_i^{S\prime} \omega_i(\hat{\psi}_m) + \Psi_m \mathbf{z}_i^{D\prime} \xi_i(\hat{\lambda}) \right) \left(\mathbf{z}_i^{S\prime} \omega_i(\hat{\psi}_m) + \Psi_m \mathbf{z}_i^{D\prime} \xi_i(\hat{\lambda}) \right)' \tag{169}$$

$$\Sigma_{1,2} = \sum_{i=1}^{n} \left(\mathbf{z}_i^{S\prime} \omega_i(\hat{\psi}_1) + \Psi_1 \mathbf{z}_i^{D\prime} \xi_i(\hat{\lambda}) \right) \left(\mathbf{z}_i^{S\prime} \omega_i(\hat{\psi}_2) + \Psi_2 \mathbf{z}_i^{D\prime} \xi_i(\hat{\lambda}) \right)' \tag{170}$$

$$\Psi_m = G_m^D \left(H^{D\prime} W^D H^D \right)^{-1} H^{D\prime} W^D \tag{171}$$

$$G_m^D = \nabla_\lambda g_m(\hat{\psi}_m) \tag{172}$$

$$H^D = \nabla_\lambda h(\hat{\lambda}) \tag{173}$$

where $\mathbf{z}^D$ and $\mathbf{z}^S$ are respectively demand and supply instruments, and W^D is the weight matrix used in the demand estimate. A similar adjustment can be made when the weight matrix is $W = (\mathbf{z}^{S\prime}\mathbf{z}^S)^{-1}$. See Hall and Pelletier (2011) for details.

B.2 EB Test

As shown in Miller and Weinberg (2017) appendix D, a consistent estimator for the asymptotic variance of $\hat{\theta}$ that takes into account the uncertainty from the demand estimate $\hat{\lambda}$ is:

$$(G^{S\prime} W G^S)^{-1} G^{S\prime} W \Sigma^{EB} W G^S (G^{S\prime} W G^S)^{-1}$$

where $G^S = \nabla_\theta g(\hat{\psi})$ and Σ^{EB} definition is analogous to Equation (169):

$$\Sigma^{EB} = \sum_{i=1}^{n} \left(z_i^{S\prime} \omega_i(\hat{\psi}) + \Psi z_i^{D\prime} \xi_i(\hat{\lambda}) \right) \left(z_i^{S\prime} \omega_i(\hat{\psi}) + \Psi z_i^{D\prime} \xi_i(\hat{\lambda}) \right)'$$

$$\Psi = G^D \left(H^{D\prime} W^D H^D \right)^{-1} H^{D\prime} W^D$$

$$G^D = \nabla_\lambda g(\hat{\psi})$$

$$H^D = \nabla_\lambda h(\hat{\lambda})$$

where $\hat{\psi} = (\hat{\theta}, \hat{\lambda})$.

B.3 AR Test

Analogous to the EB test correction, except that instead of $g(\hat{\psi})$ the supply side moment is now $z^{S\prime}(p - \hat{\Delta}_m(\hat{\lambda}) - z^S \hat{\pi}_m)$ and we take derivatives with respect to π_m.

B.4 Cox Test

Define $g_3(\hat{\psi}_1, \hat{\psi}_2) \equiv g_1(\hat{\psi}_1) - g_2(\hat{\psi}_1)$ and $Q_3(\hat{\psi}_1, \hat{\psi}_2) \equiv g_3(\hat{\psi}_1, \hat{\psi}_2)' I g_3(\hat{\psi}_1, \hat{\psi}_2)$. For ease of notation, we drop the dependency on $\hat{\psi}$ from now on. Analogous to what we show for the RV test, we can show that a consistent estimator for the limiting variance of $\sqrt{n}(Q_1 - Q_2 - Q_3)$ is:

$$\hat{\sigma}^2 = 4g_1'\Sigma_1 g_1 + 4g_2'\Sigma_2 g_2 + 4g_3'\Sigma_3 g_3 - 8g_1'\Sigma_{1,2}g_2 - 8g_1'\Sigma_{1,3}g_3 + 8g_2'\Sigma_{2,3}g_3$$

where:

$$\begin{aligned}
\Sigma_{m,3} &= \sum_{i=1}^{n} \left(z_i^{S\prime}\omega_i(\hat{\psi}_1) + \Psi_1 z_i^{D\prime}\xi_i(\hat{\lambda}) \right) \left(z_i^{S\prime}(\omega_i(\hat{\psi}_1) - \omega_i(\hat{\psi}_2)) + \Psi_3 z_i^{D\prime}\xi_i(\hat{\lambda}) \right)' \\
\Psi_m &= G_m^D \left(H^{D\prime} W^D H^D \right)^{-1} H^{D\prime} W^D \\
G_m^D &= \nabla_\lambda g_m \\
H^D &= \nabla_\lambda h(\hat{\lambda}).
\end{aligned}$$

B.5 Moments' Jacobians

To correct the supply-side standard errors for the uncertainty from the demand estimate we need the Jacobians of the supply-side and demand-side moments with respect to demand parameters. We can write them as:

$$\begin{aligned}
G^D &= \nabla_\lambda g(\lambda, \tau) = n^{-1} \sum_i -z_i^S \, \nabla_\lambda \eta_i' \\
H^D &= \nabla_\lambda h(\lambda) = n^{-1} \sum_i z_i^D \begin{bmatrix} -x_i & -p_i & \dfrac{\partial \delta}{\partial \sigma} \end{bmatrix}
\end{aligned}$$

where $\eta = \Delta^{-1}s$ and $\Delta = -O \odot \frac{\partial s}{\partial p}$.

The term $\frac{\partial \delta}{\partial \sigma}$ is straightforward to compute using the implicit function solution:

$$\frac{\partial \delta}{\partial \sigma} = - \left[\frac{\partial \tilde{s}}{\partial \delta} \right]^{-1} \left[\frac{\partial \tilde{s}}{\partial \sigma} \right]$$

For the markups' derivative we know that, for a given demand parameter λ_l:

$$\frac{\partial \eta}{\partial \lambda_l} = -\Delta^{-1} \frac{\partial \Delta}{\partial \lambda_l} \eta$$

To compute $\frac{\partial \Delta}{\partial \lambda_l}$ we use the fact that:

$$\frac{\partial s_j}{\partial p_k} = \int (\alpha + \sigma_p \upsilon_{ip}) \frac{\partial P_{ij}}{\partial \delta_k} \, dF(i) \Rightarrow$$

$$\frac{\partial s_j}{\partial p_k \partial \lambda_l} = \int \frac{\partial(\alpha + \sigma_p \upsilon_{ip})}{\partial \lambda_l} \frac{\partial P_{ij}}{\partial \delta_k} \, dF(i) + \int (\alpha + \sigma_p \upsilon_{ip}) \frac{\partial P_{ij}}{\partial \delta_k \partial \delta} \left(\frac{\partial \delta}{\partial \lambda_l} + \frac{\partial \mu_i}{\partial \lambda_l} \right) dF(i)$$

where $P_{i,j}$ is the choice probability for consumer i and product j. Note that, for $\lambda_l = \beta$ this derivative is zero, for $\lambda_l = \alpha$ we only have the first term in the right-hand side, and for $\lambda_l = \sigma \neq \sigma_p$ we only have the second term in the right-hand side. Since we already have $\frac{\partial \delta}{\partial \lambda_l}$ from matrix H^D construction, we are only missing $\frac{\partial P_{ij}}{\partial \delta_k \partial \delta}$. From Dube, Fox and Su (2012) appendix we know that for $k \neq j \neq l \neq k$:

$$\begin{aligned}
\frac{\partial^2 s_j}{\partial^2 \delta_j} &= \int P_{ij}(1 - P_{ij})(1 - 2P_{ij}) dF(i) \\
\frac{\partial^2 s_j}{\partial \delta_j \partial \delta_k} &= -\int P_{ij} P_{ik}(1 - 2P_{ij}) dF(i) \\
\frac{\partial^2 s_k}{\partial \delta_j \partial \delta_k} &= -\int P_{ij} P_{ik}(1 - 2P_{ik}) dF(i) \\
\frac{\partial^2 s_j}{\partial^2 \delta_k} &= -\int P_{ij} P_{ik}(1 - 2P_{ik}) dF(i) \\
\frac{\partial^2 s_j}{\partial \delta_k \partial \delta_l} &= \int 2 P_{ij} P_{ik} P_{il} dF(i)
\end{aligned}$$