Heteroskedastic Supply and Demand Estimation: Analysis and Testing

Matthew Grant and Anson Soderbery*

June 2023

Abstract

The Feenstra (1994) method is widely used in the international trade literature to estimate supply and demand elasticities. The method is mechanically an IV strategy, and we demonstrate that this has important implications for its application and reliability. The assumptions needed for it to yield unbiased estimates are stronger than previously understood, and in practice, estimates are subject to bias due to both weak instruments and violations of the exclusion restriction. We illustrate how these arise in context and show that standard tests identify estimates that are likely to be biased. In an application to U.S. import data, estimates of import demand and export supply elasticities are substantially lower among goods that pass standard tests relative to those that fail. We find evidence that this difference in elasticities reflects reduction of bias as well as some selection in the set of goods that tend to pass both tests.

Keywords: Import demand, Export supply, Heteroskedastic estimation
JEL Classification: F12

*Grant: Department of Economics, Dartmouth College, matthew.w.grant@dartmouth.edu. Soderbery: Department of Economics, Krannert School of Management, Purdue University, asoderbe@purdue.edu. This paper has benefited immensely from discussions with Andy Bernard, David Hummels, Rob Feenstra, Nina Pavcnik, Bob Staiger, Meredith Startz, Richard Startz, and seminar participants at Dartmouth College, Purdue University, Midwest International Trade Meetings, and Feenstra's Festrschrift. We thank Rob Feenstra and John Romalis for sharing replication data for Feenstra and Romalis (2014). Incomparable research assistance was provided by Mason Reasner. All errors and omissions are our own.

1 Introduction

The import demand elasticity is arguably the most important parameter in international economics. This elasticity of substitution is key to a wide range of results including the gains from trade (e.g., Arkolakis et al. (2012)) and microeconomic and macroeconomic adjustments to shocks. At least 50 years of research has been devoted to estimating this parameter (Hillberry and Hummels (2013)). Of course, the challenges associated with estimating demand elasticities are not unique to international economics: any estimator of demand using observed prices and quantities must navigate standard endogeneity concerns about the simultaneity of supply and demand.

One method for overcoming these challenges and estimating trade elasticities follows Feenstra (1994). This paper pioneered a model and methodology that leverages heteroskedasticity in supply and demand shocks across countries to point identify import demand and export supply elasticities. The methodology is extremely flexible, particularly in settings with many countries and many products, and does not involve tailoring instruments to different goods or time periods. It requires no information beyond prices and quantities from three or more exporters to a particular destination over time. Consequently, we have seen a wide array of applications ranging from estimating gains from product variety (c.f., Broda and Weinstein (2006)) to optimal firm scope (c.f., Hottman et al. (2016)) to optimal trade policy (c.f., Broda et al. (2008), Ossa (2014), Soderbery (2018), and Grant (2020)) and macroeconomic applications (e.g., Rigobon (2003)).

The Feenstra (1994) method relies on two core assumptions: first that supply and demand shocks are heteroskedastic across exporters, and second that supply and demand shocks are uncorrelated within an exporter.[1] To our knowledge, no prior work examines whether these assumptions are valid in practice.[2] In this paper, we expand on the insight that the Feenstra (1994) method is effectively an instrumental variables (IV) strategy, and demonstrate that this has important implications for assessing the application and reliability of the method. We show that the assumptions needed for it to yield unbiased estimates are stronger than previously understood, and that in practice, estimates may be subject to bias due to both weak instruments problems and violations of the exclusion restriction. We illustrate how these problems arise in the context of the method and show how standard tests from the applied econometrics literature can be harnessed to identify estimates that are likely to be biased.

[1] Technically, these shocks must only be uncorrelated in time and reference-country differences.

[2] The literature has pursued many refinements to Feenstra (1994) to better answer particular or general research questions; e.g., Broda and Weinstein (2006), Soderbery (2015), and Ferguson and Smith (2022) all aim to improve the properties of point estimates. However, none of these approaches have examined or addressed the fundamental question about consistency of the methodology, i.e., whether import data have sufficient heteroskedasticity and uncorrelated supply and demand shocks.

Using Monte Carlo analysis, we show that restricting attention to estimates that pass tests for weak instruments and exclusion restriction violations can substantially reduce median bias in the estimator – in some cases more than halving the bias. Applying this approach to U.S. import data, we find that elasticity estimates for many goods fail one or both tests in practice. Those that pass both generally feature lower elasticities of import demand and higher elasticities of export supply. We then discuss the implications of this bias in applications of the method, and propose several practical strategies for dealing with it.

The starting point for our paper is that the identification conditions from Feenstra (1994) are isomorphic to the requirements for an instrumental variables (IV) strategy, building on an insight first stated in the original Feenstra (1994) that one way heteroskedastic identification can be implemented using country dummies as instruments. This links the heteroskedasticity assumption from Feenstra (1994) to the standard IV estimation requirement that the instrument be correlated with the endogenous regressor. Expressing the condition in this way highlights what was previously an underappreciated aspect of heteroskedastic estimation: the risk of weak instruments. Thus, the heteroskedasticity assumption needs to be strengthened – there must be sufficient heteroskedasticity to ensure a strong first stage. Second, the Feenstra (1994) assumption that supply and demand shocks are uncorrelated is equivalent to the standard IV exclusion restriction.

Making these analogies permits us to harness tools and intuition from the IV literature in order to test the assumptions in practice. Having sufficient heteroskedasticity – i.e., having strong instruments – permits us to use results from the large literature on weak instruments including Nelson and Startz (1990); Staiger and Stock (1997), and Stock and Yogo (2005) (a nice summary is provided by Andrews et al. (2019)), and boils down to a first-stage F-test. Additionally, we show how to apply a Sargan (1958) overidentification test to test the validity of the exclusion restriction.

We demonstrate that implementing these statistical tests and restricting attention to the subset of estimates that pass both substantially improves the performance of the estimator in a set of four Monte Carlo analyses. In each analysis, we construct simulated price and expenditure share data for many suppliers of a single product. We vary the heteroskedasticity in the relevant shocks from simulation to simulation. Concurrently, in some simulations we draw the supply and demand shocks independently, while in others we intentionally create correlations between them to vary the degree and impact of violating the structural assumptions. As expected, estimates pass first-stage F-tests more often when there is more heteroskedasticity, and pass Sargan's J-test more often when there less correlation between supply and demand shocks. The median bias of the estimator is much lower among the set of realizations that pass the F- and J-tests relative to those that do not.

We then turn to a dataset of imports to the U.S. at the SITC 4-digit level from 1984-2011.[3] We evaluate the performance of the estimator in these data and find that around one half of the elasticity estimates pass both tests. Estimates that pass the test are systematically different than those that do not: they have a lower median demand elasticity and a lower median supply elasticity. These patterns are largely robust to disaggregating across industries. Looking at more and less differentiated goods based on the Rauch (1999) Classification, we find that differentiated goods are less likely to pass both tests, and that those that do pass the tests conform better to our priors about the relative elasticities of more and less differentiated products. Overall, we find evidence that there is a role for both bias reduction and selection of different types of goods when comparing elasticities for goods that do and do not pass the tests.

Finally, we discuss what can be done with estimates that appear likely to be biased. For F-test failures, one solution would be to compute weak instrument robust standard errors, although the greater uncertainty these capture would also need to be taken into account in calculations that follow from elasticity estimates. For both F- and J-test failures, a natural strategy would be to consider extensions of the basic model that capture additional mechanisms that could lead to exclusion restriction violations or weak instruments problems in the core Feenstra (1994). We focus on the endogenous quality extension developed in Feenstra and Romalis (2014),[4] but conclude that no single model is likely to solve these problems for every good, and propose that the same tests can be applied to better understand which models yield reliable estimates for which types of goods.

Although the implementation of Feenstra (1994) could be seen as an IV was first noted in the appendix of the original manuscript, the literature that followed from this seminal paper has not generally pursued this insight. Two notable exceptions are Soderbery (2010), which recognized the potential for weak instrument problems arising from finite (small) samples of trade data, and Soderbery (2015), which introduced an estimation method based on a LIML estimator to ameliorate them. Our contribution is to clarify and highlight the implications of the analogy to IV, introduce the standard test for weak instruments to the Feenstra (1994) method and show that weak instruments and exclusion restriction violations are likely introducing important biases into the huge set of empirical applications of the Feenstra (1994) method. By building a link to the econometrics literature on IV estimation, we show that the empirical international economics literature can and should harness well-understood, easily implemented tools to improve the performance of its estimation procedures.

[3] Kindly shared with us by Rob Feenstra and John Romalis; it is the data underlying Feenstra and Romalis (2014).

[4] Other models that broadly fit the discussion include Feenstra and Weinstein (2017) (endogenous markups with translog preferences), Feenstra et al. (2018) (nested demand systems) and Farrokhi and Soderbery (2022) (trade elasticities in general equilibrium). All of these papers develop heteroskedastic estimators taking into account more complicated economic structures.

The paper proceeds as follows. First, Section 2 develops the economic intuition of the estimator, showing the deep parallels (theoretically and in the data) between Leamer (1981) and Feenstra (1994). Second, Section 3 illustrates the analogy between the method and an IV, highlighting the potential problems that arise and how standard tests from the IV literature can be applied. Third, Section 4 implements a Monte Carlo analysis to understand the performance of and validate the econometric tests. Finally, in Section 5 analyze our estimates for the actual data conditional on these tests and discuss when and where to pursue alternative estimation strategies and/or modeling assumptions.

2 The Feenstra (1994) method

We begin by providing a review of the Feenstra (1994) method, with particular attention to clarifying the assumptions necessary for the method to yield unbiased elasticity estimates. We highlight why and how heteroskedasticity in supply and demand shocks drives identification, in order to build intuition for an IV analogy.

2.1 Review of the Feenstra (1994) method

Simultaneity of supply and demand is a classic problem in demand estimation: the observed relationship between price and quantity reflects an unknown combination of supply and demand forces, which prevents consistent estimation of either supply or demand elasticities from a simple regression of quantity on price. The Feenstra (1994) method solves this problem using heteroskedasticity as opposed to constructing exogenous instruments for supply and demand shocks. The flexibility of the estimator has led to broad adoption in the international economics literature, where there is frequently a need to estimate a great many elasticities (and finding sufficient instruments would be exceptionally difficult, if not impossible), and where the standard data for the field are (by design) compatible with the method. The method requires only a panel of prices and total expenditure on goods from many supplying countries over time.

To obtain this estimator, Feenstra (1994) starts with equations for demand following a constant elasticity of substitution utility function across different varieties of a good (where variety is determined by supplying country as in Armington (1969)) and supply following a "reduced form" foreign export supply function for each variety:[5]

[5] One notable alternative highlighting the flexibility of the methodology is Feenstra and Weinstein (2017), which extends this framework to a translog utility function in which the elasticities of supply and demand are varying.

$$
\begin{aligned}
S_{it} &= B_{it}\left(\frac{P_{it}}{\phi_t}\right)^{1-\sigma} && \text{(Demand)}\\
S_{it} &= e^{-\frac{A_{it}}{\omega}} P_{it}^{\frac{1+\omega}{\omega}} E_t && \text{(Supply)}, \qquad (1)
\end{aligned}
$$

where S_{it} denotes expenditure share on variety i at time t, P_{it} is the price (i.e., unit value) of the variety.[6] B_{it} is an unobserved taste parameter, ϕ_t is the CES price index accounting for all varieties, $e^{-A_{it}/\omega}$ is the exponential of a supply shock A_{it}.[7] E_t is the aggregate expenditure on all varieties, and σ is the elasticity of substitution across varieties and ω is the inverse export supply elasticity, where both elasticities are constant and shared by all varieties.[8]

Feenstra (1994) then obtains the estimating equation in three steps. First, prices and expenditure shares must be transformed to eliminate a mechanical correlation of supply and demand residuals arising from the price index. Second, supply and demand relationships are combined and averaged at the supplier level to yield an unbiased relationship between the variance in prices, the variance in shares, and the covariance between shares and prices. And third, using the relationship from the second step, Feenstra (1994) regresses the variance in price on the variance in expenditure share and covariance in price and share across suppliers to obtain unbiased estimates of supply and demand elasticities.

The first step in the method is to transform supply and demand equations to avoid a mechanical

[6] As discussed in Feenstra (1994), we rely on import expenditure shares rather than quantities in order to alleviate measurment error associated with unit value price data. Ferguson and Smith (2022) discuss potential costs of relying on market shares in place of quantities. Here, we opt to follow Feenstra (1994) for consistency.

[7] Since country i supplies less for a given P_{it} and E_t when A_{it} is large, we can interpret A_{it} as an inverse productivity shock. Exponentiating the supply shock is simply to be consistent with Feenstra (1994), and does not affect the analysis.

[8] Formally, the utility function which microfounds the demand equation is standard CES demand in which the payout from consuming a given good in a given period t, X_t, when the set of suppliers is given by $\mathbf{I}_t$ and the quantity from each supplier is denoted X_{it} is defined by

$$
X_t = \left(\sum_{i\in\mathbf{I}_t} B_{it}^{\frac{1}{\sigma}} X_{it}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}
$$

shares

$$
S_{it} = \frac{P_{it}X_{it}}{\sum_{i\in\mathbf{I}_t} P_{it}X_{it}}
$$

and the price index is given by

$$
\phi_t = \left(\sum_{i\in\mathbf{I}_t} B_{it}P_{it}^{1-\sigma}\right)^{\frac{1}{1-\sigma}}
$$

correlation in the supply and demand residuals arising from price indexes.[9] For instance, given the assumption of a CES demand system, when some supplying countries are "large" in the market (i.e., have non-zero market share) a positive supply shock from a given supplier will translate into a lower price index. Without correcting for the CES nature of demand (i.e., assuming partial equilibrium), this would create an immediate mechanical correlation in the supply and demand residuals: a positive supply shock would be correlated with a negative demand shock.[10] Feenstra (1994) corrects this problem by considering supply and demand shocks in log double differences: both with respect to a reference country and with respect to time (i.e., the previous period).[11] The transformed system is rewritten as,

$$\begin{aligned} \Delta^k s_{it} &= (1-\sigma)\,\Delta^k p_{it} + \Delta^k b_{it} && \text{(Demand)} \\ \Delta^k s_{it} &= \frac{1+\omega}{\omega}\Delta^k p_{it} - \frac{\Delta^k A_{it}}{\omega}, && \text{(Supply)} \end{aligned} \tag{2}$$

where lower case letters denote the natural log of the corresponding lowercase variable and Δ^k denotes time and reference-country differences of a variable.[12]

The second step in the method is to obtain an unbiased relationship between the variance of prices, the variance of expenditure shares, and the covariance in prices and expenditure shares at the supplier level. To do this, Feenstra (1994) starts by re-arranging the supply and demand equations, multiplying them together, and rearranging again to yield an equation of the form

$$\left(\Delta^k p_{it}\right)^2 = \tfrac{\omega}{(1+\omega)(\sigma-1)}\left(\Delta^k s_{it}\right)^2 + \tfrac{\omega(\sigma-2)-1}{(1+\omega)(\sigma-1)}\Delta^k s_{it}\Delta^k p_{it} + u_{it}, \tag{3}$$

[9] Only P_{it} and S_{it} are observed in the demand equation, thus, without correction the demand residual reflects $^{B_{it}}/_{\phi_t^{1-\sigma}}$.

[10] A lower price index will lead to lower demand from the source experiencing a "good" productivity shock. Farrokhi and Soderbery (2022) extend this logic to the supplier level. We focus on the implementation of the original Feenstra (1994) framework here, although our intuition would readily extend to the setting of Farrokhi and Soderbery (2022) as their methodology nests Feenstra (1994), and demonstrates Feenstra (1994)'s model is effectively and Eaton and Kortum (2002) model with factor mobility frictions.

[11] Although the choice of the reference country is in principle arbitrary, in practice preference is given to countries which export in as many years as possible to reduce the number of years of data which are dropped and which have a relatively large volume of exports to reduce the influence of measurement error.

[12] Formally:

$$\begin{aligned} \Delta^k s_{it} &\equiv \left[\ln\left(S_{it}\right) - \ln\left(S_{i(t-1)}\right)\right] - \left[\ln\left(S_{kt}\right) - \ln\left(S_{k(t-1)}\right)\right] \\ \Delta^k p_{it} &\equiv \left[\ln\left(P_{it}\right) - \ln\left(P_{i(t-1)}\right)\right] - \left[\ln\left(P_{kt}\right) - \ln\left(P_{k(t-1)}\right)\right] \\ \Delta^k b_{it} &\equiv \left[\ln\left(B_{it}\right) - \ln\left(B_{i(t-1)}\right)\right] - \left[\ln\left(B_{kt}\right) - \ln\left(B_{k(t-1)}\right)\right] \\ \Delta^k A_{it} &\equiv \left[A_{it} - A_{i(t-1)}\right] - \left[A_{kt} - A_{k(t-1)}\right]. \end{aligned}$$

where $u_{it} \equiv \frac{1+\omega\sigma}{(1+\omega)^2(\sigma-1)}\Delta^k b_{it}\Delta^k A_{it}$. This expression is still biased, as the supply and demand shocks are embedded in u_{it} and thus correlated with $\Delta^k p_{it}$ and $\Delta^k s_{it}$. However, the method uses the assumption that the supply and demand shocks are uncorrelated with each other at the supplier level, such that taking time averages within a supplier yields the unbiased relationship:[13].

$$\overline{(\Delta^k p_{it})^2} = \theta_1 \overline{(\Delta^k s_{it})^2} + \theta_2 \overline{\Delta^k s_{it}\Delta^k p_{it}} + \overline{u}_{it}, \tag{4}$$

where $\theta_1 \equiv \frac{\omega}{(1+\omega)(\sigma-1)}$ and $\theta_2 \equiv \frac{\omega(\sigma-2)-1}{(1+\omega)(\sigma-1)}$. The time differencing of all of the variables will remove the constant average, so that $\overline{u}_{it}$ will be the covariance of transformed supply and demand shocks and is zero in expectation (by assumption), $\overline{(\Delta^k p_{it})^2}$ will be the variance in transformed prices, $\overline{(\Delta^k s_{it})^2}$ will be the variance in transformed shares, and $\overline{\Delta^k s_{it}\Delta^k p_{it}}$ will be the covariance in transformed prices and shares.

The third step in the method is to regress variances in price on the variances in share and covariances in price and share across suppliers to obtain unbiased estimates of supply and demand elasticities. If there is heteroskedasticity in the shocks, then for at least two suppliers, Equation (4) will not be colinear. This means that the regression,

$$\widehat{\mathrm{Var}}[\Delta^k p_{it}] = \theta_1 \widehat{\mathrm{Var}}[\Delta^k s_{it}] + \theta_2 \widehat{\mathrm{Cov}}[\Delta^k s_{it}\Delta^k p_{it}] + \overline{u}_{it}, \tag{5}$$

will identify θ_1 and θ_2. Performed across suppliers, Equation 5 will yield unbiased and efficient estimates of θ_1 and θ_2, when $\widehat{\mathrm{Var}}$ and $\widehat{\mathrm{Cov}}$ denote the sample variances and covariances.[14] Finally, Feenstra (1994) shows how to rearrange the coefficient estimates, $\theta_1 \equiv \frac{\omega}{(1+\omega)(\sigma-1)}$ and $\theta_2 \equiv \frac{\omega(\sigma-2)-1}{(1+\omega)(\sigma-1)}$, to deliver consistent estimates of σ and ω, provided two key assumptions regarding supply and demand shocks hold. First, shocks are heteroskedastic. Explicitly, there need to be at least two different source countries, i and j such that $\frac{\mathrm{Var}\left[\Delta^k b_{it}\right]}{\mathrm{Var}\left[\Delta^k A_{it}\right]} \neq \frac{\mathrm{Var}\left[\Delta^k b_{jt}\right]}{\mathrm{Var}\left[\Delta^k A_{jt}\right]}$. Second, that the supply and demand shocks are uncorrelated for all source countries, i, such that $\mathbb{E}\left[\Delta^k b_{it}\Delta^k A_{it}\right] = 0$ delivers $\mathbb{E}\left[\overline{u}_{it}\right] = 0$.

[13] When conducting this regression, it is also standard (following Feenstra (1994)) to include a constant term in order to absorb covariance in the errors which is shared across suppliers (e.g., arising from measurement error in unit values)

[14] In order to make the regression efficient, Feenstra (1994) employs two-stage algorithm: the first stage regression yields consistent estimates of θ_1 and θ_2, and the second stage re-weights by the errors using the first stage coefficients to give efficient estimates.

2.2 Why Heteroskedasticity Matters: Feenstra (1994) Through the Lens of Leamer (1981)

To gain greater intuition for the Feenstra (1994) method, we turn to its theoretical foundation: Leamer (1981). The estimating Equation (4) is structural in nature, and loses some appeal in terms of economic intuition. Put another way, we do not have well established priors relating price and quantity variances and covariances. Clawing back economic intuition is best done appealing to Leamer (1981). In brief, if we consider a single exporter, Leamer (1981) shows given the system from Equation 2 we can at best set identify supply and demand elasticities. Given the variation of prices and quantity in the data, he demonstrates the true supply and demand elasticities must lie on a hyperbola given by:

$$\left((1-\hat{\sigma}) - \frac{\widehat{\mathrm{Cov}}_i[\Delta^k s_{it}, \Delta^k p_{it}]}{\widehat{\mathrm{Var}}_i[\Delta^k p_{it}]}\right)\left(\frac{1+\hat{\omega}}{\hat{\omega}} - \frac{\widehat{\mathrm{Cov}}_i[\Delta^k s_{it}, \Delta^k p_{it}]}{\widehat{\mathrm{Var}}_i[\Delta^k p_{it}]}\right) = \left(\frac{\widehat{\mathrm{Cov}}_i[\Delta^k s_{it}, \Delta^k p_{it}]^2}{\widehat{\mathrm{Var}}_i[\Delta^k p_{it}]} - 1\right)\left(\frac{\widehat{\mathrm{Cov}}_i[\Delta^k s_{it}, \Delta^k p_{it}]}{\widehat{\mathrm{Var}}_i[\Delta^k p_{it}]}\frac{\widehat{\mathrm{Var}}_i[\Delta^k s_{it}]}{\widehat{\mathrm{Cov}}_i[\Delta^k s_{it}, \Delta^k p_{it}]}\right) \tag{6}$$

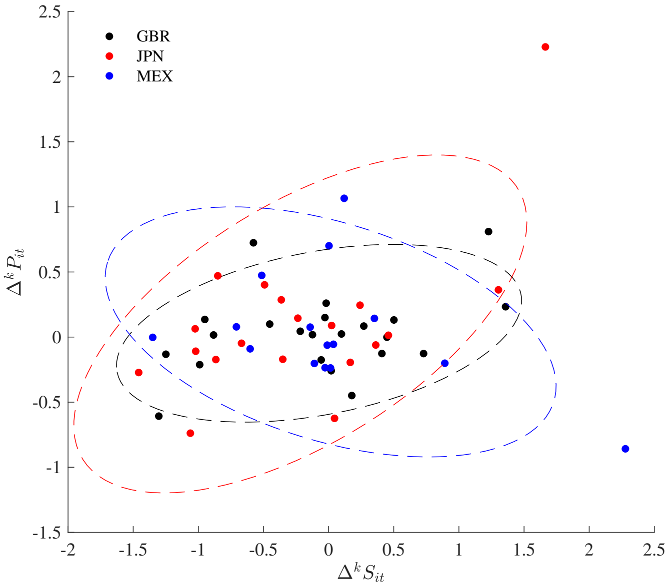
We find it useful to visualize Leamer (1981) through actual data. Consider a subset of observed data for US imports of a particular good. Figure 1 presents the analysis for the largest three exporters of wood glue (SITC 5837) to the US (Japan, Mexico and the UK). Panel (a) is the cloud of transformed data and an ellipse encasing the data.[15] Panel (b) presents the resulting hyperbola. Leamer (1981) demonstrates that if we narrow focus to the fourth quadrant of Panel (b) (i.e., where demand slopes down and supply slopes up), the variation in the data generating hyperbola can be combined to bound one elasticity but not the other. For example, UK (labeled GBR) imports suggest the true export supply elasticity satisfies $\frac{1+\hat{\omega}}{\hat{\omega}} \in [0.8, 6.2]$, and provides no information for the elasticitiy of substitution $\hat{\sigma} \in (1, \infty)$.[16] That is to say, OLS regressions of price on quantity cannot reliably point identify either supply or demand elasticities, but can be used to set identify both elasticities.

Different from Leamer (1981), Feenstra (1994) broadens the analysis to multiple hyperbola, one corresponding to each variety i. Figure 1b overlays the three resulting hyperbolae. This is where Feenstra (1994)'s innovation begins to take shape. Leamer (1981) tells us that the true estimates of supply and demand must lie on each hyperbolae. If we had only two hyperbolae, Feenstra (1994) asserts that the intersection of the hyperbolae in the feasible region is thus the point estimate for

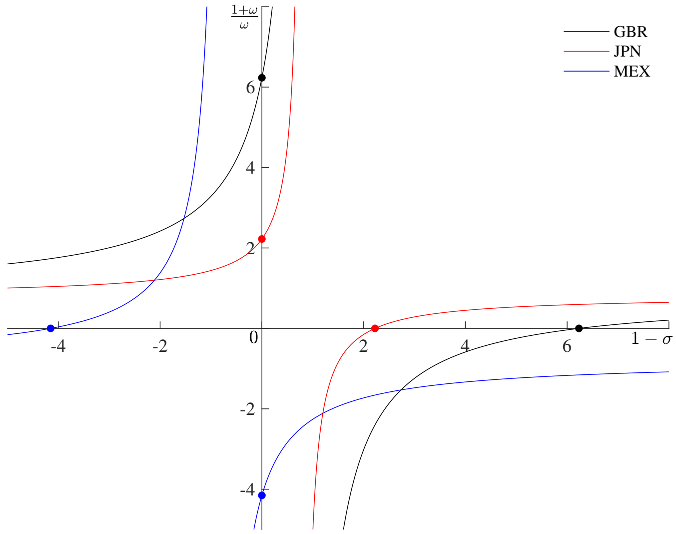
[15] The ellipse is the iso-contour of the normal distribution and represents the 90% confidence interval of the data. Informally, it is related to the R^2 of the simple regressions, and embodies the central tendency of the data. Formally, the ellipse is constructed by taking the eigen decomposition of the covariance matrix, where the eigenvectors capture the direction of greatest variance, and the eigenvalues represent the scale of the variance.

[16] These numbers differ from our extended discussion of Leamer (1981) in Appendix I as we have transformed the data following Feenstra (1994).

Figure 1: Exports of SITC 5837



(a) Clouds of Data



(b) Hyperbolae

both supply and demand given the variation in the data. In general, we can expect more than two hyperbolae for any good intersecting at potentially many points, and thus need an estimator to minimize the distance between hyperbolae. For this intuition to hold there must be sufficient heteroskedasticity in supply and demand shocks across countries (i.e., differences in the empirical hyperbolae) and the shocks must be independent (i.e., consistent hyperbolae). These are precisely the conditions for the estimator stated in Section 2.1 – $\frac{\mathrm{Var}\left[\Delta^k b_{it}\right]}{\mathrm{Var}\left[\Delta^k A_{it}\right]} \neq \frac{\mathrm{Var}\left[\Delta^k b_{jt}\right]}{\mathrm{Var}\left[\Delta^k A_{jt}\right]}$ and $\mathbb{E}\left[\Delta^k b_{it} \Delta^k A_{it}\right] = 0$.

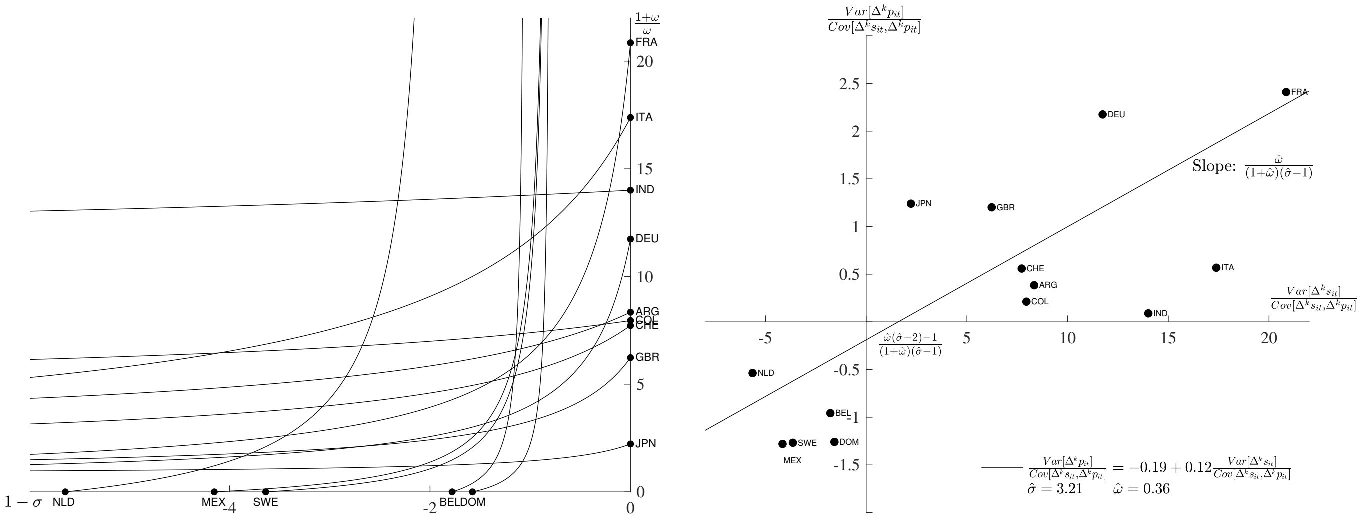
Soderbery (2015) described Feenstra (1994) as a process mapping Leamer (1981) hyperbolae to data and choosing elasticities to minimize their difference. Here, we find it beneficial to make this process explicit for both visualization and our statistical analysis to follow. First, we show that Leamer (1981) hyperbolae (Equation 6) can be arranged such that the true supply and demand elasticities must satisfy:

$$\frac{\widehat{\mathrm{Var}}_i[\Delta^k p_{it}]}{\widehat{\mathrm{Cov}}_i[\Delta^k s_{it}, \Delta^k p_{it}]} = \frac{\omega}{(1+\omega)(\sigma-1)} \frac{\widehat{\mathrm{Var}}_i[\Delta^k s_{it}]}{\widehat{\mathrm{Cov}}_i[\Delta^k s_{it}, \Delta^k p_{it}]} + \frac{\omega(\sigma-2)-1}{(1+\omega)(\sigma-1)}. \tag{7}$$

Noticing the similarities between Feenstra (1994)'s estimating equation and the representation of Leamer (1981) hyperbolae above, we can divide Equation (5) through by the $\widehat{\mathrm{Cov}}[\Delta^k s_{it}, \Delta^k p_{it}]$:

$$\frac{\widehat{\mathrm{Var}}_i[\Delta^k p_{it}]}{\widehat{\mathrm{Cov}}_i[\Delta^k s_{it}, \Delta^k p_{it}]} = \theta_1 \frac{\widehat{\mathrm{Var}}_i[\Delta^k s_{it}]}{\widehat{\mathrm{Cov}}_i[\Delta^k s_{it}, \Delta^k p_{it}]} + \theta_2 + \frac{\bar{u}_{it}}{\widehat{\mathrm{Cov}}_i[\Delta^k s_{it}, \Delta^k p_{it}]}. \tag{8}$$

Figure 2: Exports of SITC 5837



(a) Truncated hyperbolae (b) Hyperbolae as data

Now we can explicitly see how Feenstra (1994) maps exporter price and quantity data to empirical hyperbolae as data points, then chooses supply and demand elasticities that minimize their distances. Put another way, Equation 8 is identical to Equation 7 with an error term.[17] Feenstra (1994) thus extracts supply and demand information from exporter flows in the form of empirical hyperbolae, then, estimating elasticities that minimize their distance, jointly identifies both supply and demand.

Figure 2 presents the outcome of mapping each hyperbola into data and estimation for SITC 5837. Panel (a) displays all of the exporter hyperbolae for SITC 5837 constrained to the feasible region. There are 14 viable exporters of SITC 5837 to the US. Each of these exporters generate a hyperbola, which are constrained to the feasible region and plotted in Figure 2a. Here we can see the importance of mapping the information content of hyperbolae to usable data. There are many intersections of hyperbolae, and their information content requires compression. Figure 2b presents the mapping to empirical hyperbolae from Equation (8) along with the line of best fit from Feenstra (1994)'s regression.[18] The resulting supply and demand elasticities are estimated to be $\hat{\sigma} = 3.21$ and $\hat{\omega} = 0.36$ for SITC 5837. The line generated from these estimates minimizes the weighted distance between these empirical hyperbolae. Here, we see how Feenstra (1994) extracts information on supply

[17]To be clear, the Feenstra (1994) estimating equation in Equation 5 for a single supplier i is the same as the expression for a Leamer (1981) hyperbola, except that it is expressed in time and reference country differences and shares, which makes no difference for the Leamer (1981) method. Referring to Appendix I, where we discuss Leamer (1981) in more detail, after we define $\beta = 1 - \sigma$ and $\gamma = \frac{1+\omega}{\omega}$, Equation 8 is identical to Equation 11 with an error term. Note also that the assumption necessary for Leamer (1981) – uncorrelated supply and demand shocks – is also assumed in Feenstra (1994). This relationship was discussed briefly in Feenstra (1994) Footnote 6 and analyzed in greater detail by Soderbery (2015).

[18]We provide more information about our implementation in Section 4.

and demand processes from the heteroskedastic shocks faced by each country in order to pin down estimates of supply and demand.

2.2.1 Key Assumptions

As we have highlighted in our review, the Feenstra (1994) method requires two key conditions to yield unbiased estimates. First, supply and demand shocks need to be uncorrelated for all suppliers.[19] This assumption is necessary for the observed second moments of the joint distribution of price and quantity to be informative about the supply and demand elasticities. Under this condition the observed prices and quantities for a given supplier constrain the supply and demand elasticities to lie on a hyperbola.

Second, supply and demand shocks must be heteroskedastic across countries, so that the hyperbolae for different suppliers are different. If this condition fails, then all the hyperbolae lie on top of each other and the elasticities are only set identified (instead of point identified) as in Leamer (1981). Next, we show how these assumptions are analogous to standard IV requirements, and draw out the practical implications of this analogy.

3 Feenstra (1994) as an Instrumental Variables Strategy

In this section, we first highlight the isomorphism between the assumptions underlying Feenstra (1994) and the requirements for a valid IV estimation strategy. This should be unsurprising, since in practice Feenstra (1994) is implemented as a two stage least squares estimator. However, this equivalence has not typically been remembered or explored in the empirical literature applying the method.

We build on this insight, showing that the equivalence to IV implies there are important and generally unrecognized potential pitfalls facing the empirical international economics literature. This also allows us to harness the tools and intuition created by an extensive literature evaluating the performance of IV strategies to test whether the assumptions underlying Feenstra (1994) hold in practice.

[19] When a constant is included in the regression, then the requirement is that the correlation in supply and demand shocks must be the same for all suppliers.

3.1 Assumptions and Requirements for an IV

The appendix to Feenstra (1994) argues that the method can be implemented as an IV strategy, where country dummies can be used as instruments to move from Equation 3 to the estimating Equation 5. We further the discussion by showing the underlying assumptions of the estimator can be mapped to the standard assumptions for an IV estimation strategy. We will then use this mapping to motivate the relevance of statistical tests commonly used to evaluate the performance of IV methods in the context of Feenstra (1994).

3.1.1 Exclusion Restriction

The exclusion restriction in a standard IV strategy requires that the instrument not be correlated with the regressand if the endogenous regressors are held fixed (i.e., the instrument must be uncorrelated with the error).

In the Feenstra (1994) setting, the instruments are simply source country dummies. Denote these dummies with an indicator $\mathbf{1}(j)$ for country j; then the exclusion restriction requires $\mathbb{E}\left[\mathbf{1}(j)\, u_{it}\right] = 0$. Of course, this will only be true if and only if, for every source j, $\mathbb{E}\left[u_{jt}\right] = 0$.

This is exactly the assumption of uncorrelated supply and demand shocks provided by Feenstra (1994), which requires that, for every supplier, the supply and demand shock in reference differences are uncorrelated. Recall from Section 2 that $u_{it} = \frac{1+\omega\sigma}{(1+\omega)^2(\sigma-1)}\Delta^k b_{it}\Delta^k A_{it}$. Thus, if $\mathbb{E}\left[u_{it}\right] = 0$ for all sources i, this implies that supply and demand shocks are uncorrelated in time and reference differences.[20]

3.1.2 Relevance Requirement

The second requirement for a standard IV strategy is that the instrument be correlated with the variable of interest. More formally, the covariance matrix of the instruments with the endogenous regressors in the first stage of the IV regression must have full rank.

In the Feenstra (1994) setting, with source country dummies as instruments, there are two endogenous regressors,[21] and for the matrix to have full rank it must be there are countries j and l

[20]Since it is standard to run the regression to absorb shared covariance in the errors, in practice this assumption will only fail if the covariances are different across different suppliers, i.e., there are non-reference suppliers i and j such that $\mathbb{E}\left[u_{it}\right] \neq \mathbb{E}\left[u_{jt}\right]$.

[21]In specifications with a constant (as is typical in practice) there will be 3 such regressors.

such that the matrix,

$$\begin{bmatrix} \mathbb{E}\left[\mathbf{1}(j)\left(\Delta^k s_{it}\right)^2\right] & \mathbb{E}\left[\mathbf{1}(j)\,\Delta^k s_{it}\Delta^k p_{it}\right] \\ \mathbb{E}\left[\mathbf{1}(l)\left(\Delta^k s_{it}\right)^2\right] & \mathbb{E}\left[\mathbf{1}(l)\,\Delta^k s_{it}\Delta^k p_{it}\right] \end{bmatrix} = \begin{bmatrix} \mathbb{E}\left[\left(\Delta^k s_{jt}\right)^2\right] & \mathbb{E}\left[\Delta^k s_{jt}\Delta^k p_{jt}\right] \\ \mathbb{E}\left[\left(\Delta^k s_{lt}\right)^2\right] & \mathbb{E}\left[\Delta^k s_{lt}\Delta^k p_{lt}\right] \end{bmatrix},$$

has full rank. For the matrix to have full rank, it must have a non-zero determinant, i.e., that

$$0 \neq \mathbb{E}\left[\left(\Delta^k s_{jt}\right)^2\right] \cdot \mathbb{E}\left[\Delta^k s_{lt}\Delta^k p_{lt}\right] - \mathbb{E}\left[\left(\Delta^k s_{lt}\right)^2\right] \cdot \mathbb{E}\left[\Delta^k s_{jt}\Delta^k p_{jt}\right].$$

If we assume that the exclusion restriction holds for all suppliers (i.e., the supply and demand shocks are uncorrelated), then we can follow Leamer (1981) to write this as,

$$\begin{aligned} 0 \neq & \left(\frac{\left(\frac{1+\omega}{\omega}\right)^2 \mathrm{Var}\left[\Delta^k b_{jt}\right] + \left(\frac{\sigma-1}{\omega}\right)^2 \mathrm{Var}\left[\Delta^k A_{jt}\right]}{\left(\sigma+\frac{1}{\omega}\right)^2}\right) \cdot \left(\frac{\left(\frac{1+\omega}{\omega}\right) \mathrm{Var}\left[\Delta^k b_{lt}\right] + \left(\frac{\sigma-1}{\omega^2}\right) \mathrm{Var}\left[\Delta^k A_{lt}\right]}{\left(\sigma+\frac{1}{\omega}\right)^2}\right) \\ & - \left(\frac{\left(\frac{1+\omega}{\omega}\right)^2 \mathrm{Var}\left[\Delta^k b_{lt}\right] + \left(\frac{\sigma-1}{\omega}\right)^2 \mathrm{Var}\left[\Delta^k A_{lt}\right]}{\left(\sigma+\frac{1}{\omega}\right)^2}\right) \cdot \left(\frac{\left(\frac{1+\omega}{\omega}\right) \mathrm{Var}\left[\Delta^k b_{jt}\right] + \left(\frac{\sigma-1}{\omega^2}\right) \mathrm{Var}\left[\Delta^k A_{jt}\right]}{\left(\sigma+\frac{1}{\omega}\right)^2}\right), \end{aligned}$$

which, after substantial algebra, simplifies to the heteroskedasticity condition from Feenstra (1994). Specifically, that $\exists\, i \neq j$ s.t. $\frac{\mathrm{Var}\left[\Delta^k b_{jt}\right]}{\mathrm{Var}\left[\Delta^k b_{lt}\right]} \neq \frac{\mathrm{Var}\left[\Delta^k A_{jt}\right]}{\mathrm{Var}\left[\Delta^k A_{lt}\right]}$. Analogous steps suffice to show that the converse is also true – when the heteroskedasticity condition from Feenstra (1994) is satisfied then covariance matrix of the instruments with the endogenous regressors will have full rank and the IV relevance assumption is therefore satisfied.

3.2 Pitfalls and Tests

With the isomorphism between the assumptions for Feenstra (1994) and the conditions for valid IV estimation in hand, we next turn to how failures of these assumptions might cause problems for elasticity estimation, and how the validity of the assumptions can be tested in practice to avoid these pitfalls.

3.2.1 Heteroskedasticity and Weak Instruments

The relevance condition for Feenstra (1994) implemented as an IV is that there must be heteroskedasticity in the (relative) supply and demand shocks across countries. To test whether this condition holds in practice, we outline a bootstrapping procedure in the Appendix and apply it to data on goods imported into the US. Consistent with the intuition above, the heteroskedasticity condition is simply a requirement that country level Leamer (1981) are different. We resample the data with replacement to calculate the probability that at least two of the hyperbola for any given imported good do not coincide. Ultimately, we find that this is rarely a binding constraint – nearly every product we look at in the trade data passes this requirement.

Ultimately, the simple existence heteroskedasticity condition provided in Feenstra (1994) is in fact not strong enough. As shown in the prior subsection, this condition is equivalent to requiring that there be an invertible first stage. While necessary, a substantial literature following Nelson and Startz (1990) and Staiger and Stock (1997) has shown that this condition is insufficient for proper performance of an IV estimator. Instead, for proper performance of an IV-estimator, there must be very high confidence that the coefficient in the first stage is not 0, as even a very small probability that the first stage is not in fact invertible can introduce substantial bias into estimates.

It is generally understood that standard IV estimation can yield poor inference when even a few instruments are weak, even if the the problem is overidentified and the strong instruments are sufficient in number for full identification alone (see, e.g., Stock and Yogo (2005)). For many product categories, the fact that the U.S. imports from many countries implies that there are a large number of instruments for each elasticity to be estimated. This makes the Feenstra (1994) method particularly susceptible to weak instrument problems, since with more suppliers the risk of drawing at least one weak instrument increases (consistent with simulations run in Soderbery (2010)). This problem can be overcome by using a limited information maximum likelihood (LIML) approach instead of a standard IV estimation (as in Soderbery (2015)).

Fortunately, the econometrics literature also offers a way to determine whether weak instruments concerns are likely to apply. The solution is to use a first-stage F-test for weak instruments, such that estimates for which, given the number of instruments, the first stage F-statistics is below a cutoff value are more likely to suffer from bias than those which do not.[22]

3.2.2 Correlated Shocks and Endogeneity

We next turn to the exclusion restriction, which for Feenstra (1994) requires that supply and demand shocks be uncorrelated. Correlation of supply and demand shocks will introduce bias (since in this case there will be an omitted variable capturing the correlation by supplier). The literature has particularly focused on endogenous quality – in which more productive firms find it profitable to invest

[22] Note that there is a gap between the current many weak instruments literature (e.g., Andrews et al. (2019)) and heteroskedastic estimators. Current econometric methods to analyze a setting with multiple endogenous regressors (we have two), many potentially weak instruments (each exporter indicator is an instrument) and assume that heteroskedastic shocks do not exist.

more in their quality. This mechanism is modeled and corrected by Feenstra and Romalis (2014) through a modified heteroskedastic estimator. However, this is not the only possible mechanism which could create a correlation in the supply and demand residuals, e.g., misspecification of the models of demand or supply could equally create an apparent correlation in cost and taste residuals (e.g., Feenstra and Weinstein (2017), Feenstra et al. (2018), and Farrokhi and Soderbery (2022) all explore other potential types of misspecification).

Generally, applications of Feenstra (1994) and related models are overidentified: there are two parameters to be estimated and many more than two supplying countries.[23] It is well-known that in settings with more instruments than regressors, it is possible to test the validity of the exclusion restriction.

Although there are multiple methods for testing the validity of the exclusion restriction in overidentified models, the standard test is the Sargan J-test. This test can be understood as a test of internal consistency: whether the estimated residuals are orthogonal to the full set of instruments. It has been pointed out (see, e.g., Parente and Silva (2012) for full details) that passing this test is not sufficient to ensure the instruments are valid. However, this test is widely used and in settings where the test rejects the null the instruments are unlikely to all be valid.[24] We explore the details of the implementation of a J-test in a trade setting subsequently.

4 Monte Carlo Analysis

In this section, we demonstrate how F-tests and J-tests can be applied in practice to identify weak instruments problems and potential violations of the exclusion restriction. We outline our approaches to estimating elasticities using a modified Feenstra (1994) method, applying the two statistical tests, and comparing the features of estimates that do and do not pass the tests.

To evaluate the performance of the estimator and the efficacy of IV based statistical tests, we use Monte Carlo analysis. First, we show that when we construct simulated data which are broadly representative of real trade data, that the assumptions of Feenstra (1994) are sometimes violated and this yields moderate bias in our median elasticity estimates. Second, we show that the standard tests are effective at detecting violations of the identifying assumptions in the specific context of Feenstra (1994) method applied to data on traded goods. And third, we show that, conditional on

[23] Since estimation usually involves a constant and one country is used as a reference, the model is overidentified if there are at least 4 suppliers.

[24] Parente and Silva (2012) argue that the test can falsely reject when there are different local average treatment effects for different instruments. However, in the Feenstra (1994) setting the validity of the method depends on structural assumptions which require that all instruments have the same local average treatment effects, so this criticism is not valid in our context.

passing the tests, estimates have substantially less residual median bias relative to those that do not. To accomplish these goals, we perform four different Monte Carlo simulations reflecting, different violations of the identifying assumptions.

4.1 Estimation Approach

In both this and the following section of the paper, we use publicly available ComTrade data, which record unit values and quantity data for goods traded between every pair of countries in the world.[25] These data are coded at the 4-digit SITC level, and cover the period from 1984-2011.

When implementing the Feenstra (1994) method, we build on Soderbery (2015)'s nonlinear limited information maximum likelihood (LIML) routine, and extend the method to a heterosekedasticy-robust nonlinear LIML following Hausman et al. (2012). Soderbery (2015) shows that using a LIML routine can ameliorate the weak instruments problem that arises in the trade setting when there are many different source countries, weighting instruments by their performance in the first stage. Thus, source countries with relatively precisely estimated variances of price and share and covariance between price and share will have points which better fit on a hyperbola and a greater weight will be placed on these hyperbolae.[26] It is relevant to note, therefore, that all of the problems identified by the F-test we apply below are those that appear above and beyond those that are already eliminated by the use of a LIML routine – we should expect weak instruments problems to be even more prevalent using the approach that is more standard in the literature. The extension to a heterosekedasticy-robust nonlinear LIML following Hausman et al. (2012) further makes the estimator robust to heteroskedasticity in the first stage.

4.2 Implementing Simulations

In our Monte Carlo analysis, we start with a choice of "true" supply and demand elasticities and simulated supply and demand shocks which preserve key moments from estimated trade residuals. We combine these objects to obtain simulated prices and expenditure data in an artificial dataset for a single good. We then apply our modified Feenstra (1994) estimator and compare our estimated elasticities with the true values over many different draws of the supply and demand shocks. We also adjust some moments of the supply and demand shocks (while preserving other ones) in some

[25] To facilitate comparisons later on, we use the exact data described in Feenstra and Romalis (2014), including cleaning.

[26] A related point to bear in mind is that the asymptotic properties of Feenstra (1994)'s estimator are based on the number of time periods rather than the number of varieties (i.e., hyberbolae). Put simply, hyperbolae are more believable when observing many shocks over time. In contrast, adding imprecise hyperbolae does not improve the estimator.

of the simulations in order to show how different violations of the estimating assumptions affect the estimates.

Here we introduce a third parameter for evaluation, $\rho \equiv \frac{\omega(\sigma-1)}{1+\omega\sigma}$, to consider alongside the demand and supply elasticities, σ and ω. The parameter ρ captures correlation between vertical shifts in the demand curve (i.e., demand shocks) and changes in the equilibrium price, and thus conveys useful information about the overall system of supply and demand. In our simulations, we assume that $\sigma = 2.5$, $\omega = 0.5$ and $\rho = 0.43$ for all simulations, and use these assumed elasticities to translate simulated shocks into time and reference differenced prices and shares.[27] Our draws of the residuals are designed to replicate the estimated residuals of SITC 6997 (Articles of Copper, Tin, etc.).[28] We also break the estimated residuals into a common component (the average of the estimated residual across all suppliers of a given good in a given year) and an idiosyncratic component (which is the difference between the estimated residual and the common component). Then we summarize these estimated residuals (and their common and idiosyncratic components), by the average correlation coefficient between the shocks (averaged across all suppliers), and the mean and variance of the distribution of the idiosyncratic and common components of the variances of the estimated residuals across all suppliers. When taking new draws of simulated residuals, we constrain them to maintain this structure by matching some moments of the empirical residual distribution. We also match the number of suppliers, the average and standard deviation of the number of years per supplier. Full details of our Monte Carlos are provided in the Appendix.

We run four different Monte Carlos that simulate different violations of the assumptions on heteroskedasticity and shock correlation. In the first Monte Carlo, Purge Correlation (MC1), we set the correlation coefficient for supply and demand shocks to zero and set the variances of the common shocks to zero. This scenario most closely captures an ideal setting for the estimator in which the exclusion restriction is valid and the instruments are strong. For our second Monte Carlo, Shock Correlation (MC2), we set the variances of the common shocks to zero (note that we do not adjust the correlation coefficient). This scenario captures failures of the exclusion restriction but strong instruments. For our third Monte Carlo, Common Correlation (MC3), we set the correlation coefficient for supply and demand shocks to zero (note that we do not adjust the variances of the common shocks); this scenario captures weak instruments but a valid exclusion restriction. And for our fourth Monte Carlo, Common and Shock Correlation (MC4), we do not adjust our calculated moments, thus preserving both the correlation coefficient and the variances of the common shocks.

[27] These elasticities are close to the median elasticities in the dataset.

[28] This good is a reasonable representative of the dataset as a whole: it is close to the median for its estimated elasticities, the p-value of rejecting the null under the Sargan J-test, and the ratio of its F-statistic to the relevant cutoff from Stock and Yogo (2005).

This final scenario combines weak instruments and failures of the exclusion restriction.

Table 1: Monte Carlos Across Specifications

Metric		Purge Correlation (MC1)	Shock Correlation (MC2)	Common Correlation (MC3)	Common and Shock Corr (MC4)
Correlations	$\overline{\text{Corr}}[\Delta^k b_{it}, \Delta^k A_{it}]$	0.000	0.270	0.000	0.270
	$\frac{\text{Corr}[\Delta^k A_t]}{\overline{\text{Corr}}[\Delta^k A_{it}]}$	0.000	0.000	0.440	0.440
	$\frac{\text{Corr}[\Delta^k b_t]}{\overline{\text{Corr}}[\Delta^k b_{it}]}$	0.000	0.000	0.640	0.640
Median Bias	σ	−0.005	−0.127	0.000	−0.095
	ω	−0.008	0.124	−0.004	0.125
	ρ	−0.011	−0.054	−0.003	−0.021
Median Test Statistics	F^{CD}	2.005	2.038	1.869	1.952
	F^{KP}	2.311	2.220	2.168	2.085
	$J^{\text{P-value}}$	0.383	0.423	0.291	0.324

Notes: Correlations describe the correlation patterns underlying the distribution supply and demand shocks are drawn from for each Monte Carlo. Median Bias is the median across all repetitions of the relative difference between the estimated parameter and the actual parameter, e.g., Median Bias $\sigma \equiv \frac{\hat{\sigma}-2.5}{2.5}$. Median Test Statistics are the medians across repetitions, where F^{CD} and F^{KP} are Cragg-Donald and Kleibergen-Paap F-statistics and $J^{\text{P-value}}$ is the p-value of the estimated J-statistic.

Our simulated data and the degree of bias in the resulting elasticity estimates are summarized in Table 1. In the first set of rows describes the correlations in the error terms that are built into the design of each set of simulated data. The first row describes covariances in the supply and demand shock within supplier, while the second captures covariances supply shocks across suppliers in a given year (i.e., common correlated supply shocks) and the third row captures covariances in demand shocks for varieties provided by different suppliers (i.e. common correlated demand shocks).

The second set of rows capture median bias across 1000 simulations for each Monte Carlo. Median bias is calculated as the percentage difference between the true elasticity (which is held constant by design across simulations) and the median estimated elasticity (which varies across simulations due to different draws of supply and demand shocks). MC1, in which the data is closest to that envisioned by the estimator, shows low median bias. The bias is largest in MC2 and MC4 where the fundamental assumption of uncorrelated shocks is violated.

The last set of rows describes median F-stats (Cragg-Donald and Kleibergen-Papp) and the p-value of a Sargan J-test against the null that the exclusion restriction is valid. As expected, the F-stats are lower in the presence of common correlated shocks (MC3 and MC4), while the p-values are higher in the case of failures of the exclusion restriction (MC2 and MC4). Broadly, common correlated shocks tend to reduce first stage F-statistics, but seem to have little impact on median bias, as we can see by comparing MC1 with MC3 and MC2 with MC4.

By screening our estimates using the tests and restricting attention to only those that pass them, we can reduce the bias, as we summarize in Table 2. Each of the four groups of rows summarizes results for a different Monte Carlo. In the first two columns, we summarize the bias in the estimates for simulations that fail versus pass the cutoff for the F-statistic. In the next two columns, we summarize the bias in the estimates for simulations that fail vs pass the cutoff for the J-statistic. And the in final two columns, we summarize the bias in the estimates for simulations that fail either test vs those which pass the cutoff for both. We adopt as cutoffs an 80% probability of rejecting the null for the Sargan J-test, and a 25% maximal size of a 5% Wald test from Stock and Yogo (2005) for LIML estimation for the F-statistic.

Table 2: Monte Carlos across Specifications and Rules of Thumb

		F-test		J-test		Both Tests	
Scenario	Metric	Fail	Pass	Fail	Pass	Fail	Pass
MC1	Bias $\hat{\sigma}$	-0.027	0.004	-0.005	-0.005	-0.018	0.005
	Bias $\hat{\omega}$	-0.024	-0.008	0.028	-0.015	-0.012	-0.011
	Bias $\hat{\rho}$	-0.036	-0.005	0.010	-0.013	-0.021	-0.006
	F^{KP}	1.632	2.708	2.498	2.285	1.763	2.691
	$J^{\text{P-value}}$	0.359	0.464	0.880	0.359	0.544	0.379
	Share	0.323	0.677	0.149	0.851	0.439	0.561
MC2	Bias $\hat{\sigma}$	-0.169	-0.104	-0.108	-0.129	-0.153	-0.105
	Bias $\hat{\omega}$	0.180	0.186	0.220	0.177	0.192	0.178
	Bias $\hat{\rho}$	-0.109	-0.029	-0.026	-0.061	-0.084	-0.030
	F^{KP}	1.651	2.688	2.346	2.182	1.776	2.689
	$J^{\text{P-value}}$	0.391	0.510	0.893	0.368	0.598	0.405
	Share	0.366	0.634	0.169	0.831	0.489	0.511
MC3	Bias $\hat{\sigma}$	-0.022	0.011	-0.009	0.000	-0.021	0.013
	Bias $\hat{\omega}$	0.013	-0.013	0.014	-0.008	0.013	-0.014
	Bias $\hat{\rho}$	-0.018	0.002	0.004	-0.005	-0.014	0.004
	F^{KP}	1.605	2.656	2.427	2.122	1.677	2.652
	$J^{\text{P-value}}$	0.272	0.376	0.878	0.301	0.333	0.331
	Share	0.399	0.601	0.098	0.902	0.471	0.529
MC4	Bias $\hat{\sigma}$	-0.131	-0.073	-0.128	-0.093	-0.125	-0.070
	Bias $\hat{\omega}$	0.195	0.178	0.183	0.185	0.194	0.177
	Bias $\hat{\rho}$	-0.038	-0.007	-0.019	-0.021	-0.032	-0.007
	F^{KP}	1.607	2.628	2.197	2.066	1.674	2.646
	$J^{\text{P-value}}$	0.310	0.414	0.888	0.318	0.385	0.368
	Share	0.416	0.584	0.118	0.882	0.499	0.501
Rules	$J^{\text{P-value}} < 0.8$			✗	✓	✗	✓
	$F^{\text{P-value}} < 0.025$	✗	✓			✗	✓

Notes: Bias is the median across all repetitions in the described bin of the relative difference between the estimated parameter and the actual parameter, e.g., Bias $\hat{\sigma} \equiv \frac{\hat{\sigma}-2.5}{2.5}$. F^{KP} is the median Kleibergen-Paap F-statistic, $J^{\text{P-value}}$ is the p-value of the J-statistic, and Share Pass is the fraction of repetitions in the bin such that the corresponding test passes or fails our cutoff rules. Rules are cutoff rules defining Fail versus Pass.

The choice of cutoffs is a tradeoff between type I and type II error – tighter cutoffs will reduce bias, at the cost of a smaller sample that passes the test. Since the supply and demand shocks are in a difference-in-differences format (i.e., time and reference country differences), this should remove most of the correlations. And the inclusion of a constant will absorb failures of the exclusion restriction which are shared across all source countries. Thus led us to a weak prior in favor of the null and an 80% p-value.[29]

Passing the F- and J-tests substantially improves the performance of the estimates. In MC1, where there is not much bias in the full set of simulations, there is still less bias conditional on passing either test or both of them. These effects are stronger in MC2-MC4 where there is more unconditional bias. Although there is some median bias even conditional on passing both tests, it is generally moderate.

While we expect these tests to detect bias, it was not ex ante clear how good these tests would be at detecting biased estimates. In particular, overidentification tests are generally known to have low statistical power. One mitigating factor in our setting is that there are a great many more instruments (exporting countries) than endogenous regressors. Consequently, there is more power in the test than there might be in a more typical setting with less overidentification.

To summarize, our Monte Carlo analysis demonstrates that plausible violations of the identification assumptions can generate large biases in elasticities estimated using the Feenstra (1994) method in a trade data context. It also shows that – in a controlled setting where we know the "true" elasticity and therefore can calculate bias with certainty – the implementation of the F-test and J-test successfully identify estimates where the identification assumptions are violated, and allow researchers to focus attention on estimates that are less subject to bias.

5 Test Statistics and Estimates with Data

Having validated the performance of the tests in Monte Carlo simulations, we turn to applying it to estimates on real trade data. We wish to answer several questions. Do estimates for any goods pass the tests, and if so, what share? Are there systematic differences in elasticities between goods that pass the tests and those that do not? If so, can we gain any insight into how much of the difference in elasticity estimates is due to bias as opposed to some other unobservable differences

[29] Our choice of cutoff for the first-stage F-test is more difficult. Current econometric methods to analyze a setting with multiple endogenous regressors (we have two), many potentially weak instruments (each exporter indicator is an instrument) and a heteroskedastic first stage (consistent with our use of the Hausman et al. (2012) method) do not exist. Ultimately, we use cutoffs from Stock and Yogo (2005). Although these cutoffs may not have exactly the same asymptotic properties in terms of the probabilities of error of a certain size, we assert they should be correlated with first stage strength and reduce bias in the estimates.

between goods that pass and those that do not? And finally, what steps can a practitioner take in the case of estimates that fail the tests?

5.1 Applying the Tests to Data

We apply the estimator to our data on U.S. imports, and we present the results in Table 3. Although our dataset includes 735 goods, estimation is only possible for 656 of them.[30] The first column of Table 3 describes the sub-groups summarized in the following columns. There are two broad categories: all 656 goods with viable estimates together and goods subdivided based on whether they pass the F- and J-tests.

Table 3: Impact of the Weak Instruments and Sargan Tests in the Data

			Median				
Data	Subset	Products	$\hat{\sigma}$	$\hat{\omega}$	$\hat{\rho}$	F^{KP}	$J^{\text{P-value}}$
Full Data	All	656	2.668 (0.131)	0.119 (0.341)	0.061 (0.026)	8.293	0.565
Subsets Applying Rules of Thumb	Fail F	137	2.718 (0.227)	0.274 (0.726)	0.180 (0.049)	1.293	0.281
	Pass F	519	2.620 (0.115)	0.098 (0.295)	0.048 (0.021)	16.045	0.654
	Fail J	270	3.244 (0.253)	0.001 (0.787)	0.003 (0.034)	10.709	0.997
	Pass J	386	2.310 (0.085)	0.313 (0.176)	0.153 (0.020)	6.479	0.132
	Fail Both	38	4.015 (0.870)	0.001 (1.443)	0.007 (0.093)	1.404	0.986
	Pass F not J	232	3.169 (0.236)	0.001 (0.701)	0.002 (0.030)	15.913	0.998
	Pass J not F	99	2.430 (0.170)	0.408 (0.422)	0.289 (0.035)	1.207	0.084
	Fail Any	369	3.094 (0.232)	0.030 (0.713)	0.014 (0.034)	4.106	0.970
	Pass Both	287	2.206 (0.054)	0.243 (0.106)	0.121 (0.012)	16.045	0.149

Notes: Products are the count of 4-digit SITC industries. Median estimates across designated products for demand and supply are denoted by $\hat{\sigma}$, $\hat{\omega}$, and $\hat{\rho}$. The median robust standard errors from HLIML are in parentheses for our estimates of σ, ω, and ρ. F^{KP} is the median Kleibergen-Paap F-statistic, $J^{\text{P-value}}$ is the p-value of the J-statistic. Pass versus Fail follow our rules of thumb described in Table 2.

The remaining columns describe the count of goods with estimates falling into the given category along with median estimates in the given category. We provide medians rather than means as some of

[30] Recall the minimum requirements of the estimator are at least three exporters (varieties) each exporting for three periods (years).

the estimated elasticities can be very high or low -- an elasticity of 1000 is not substantively different than an elasticity of 100 (both describe nearly perfect elastic supply or demand) but the former has a much larger effect on means.[31] Medians are robust to this issue.

A glance at Table 3 answers two of our questions. First, a substantial share of estimates (nearly 45% of those for which estimation is feasible) pass both tests. And second, the table shows a general pattern that estimates which pass the weak instruments and Sargan tests feature lower elasticities of demand and lower elasticities of supply (i.e., larger estimated omegas, since omega is the inverse elasticity of supply). Roughly speaking, the first-stage F-test is more closely tied to the export supply elasticity: passing the F-test accompanies a lower ω and thus more elastic supply. And the J-test is more closely tied to the demand elasticity: passing the J-test accompanies a lower σ and thus more inelastic demand.

Overall, estimates for goods which pass both tests have estimates of $\hat{\sigma}$ which are roughly 25% lower than the overall sample (2.206 vs 2.668) and estimates of foreign inverse export supply elasticity are roughly twice as large as the overall sample (0.243 vs 0.119). Finally, we present the impact of passing the tests on the median $\hat{\rho}$. Passing the first stage F-test is associated with a lower estimate of $\hat{\rho}$, while passing Sargan's J-test has the opposite effect. On net, goods that pass both tests have estimates of $\hat{\rho}$ which are roughly double.[32]

5.2 Composition vs Bias

Differences in estimated elasticities for goods that pass F- and J-tests versus those that do not could be due to either the elimination of bias or due to differences in the types of goods that pass (e.g., if having different true elasticities is correlated with the distribution of shocks which makes the estimates more likely to pass one or both of the tests). Although we cannot provide a complete answer to this question – we do not have reliable estimates of the elasticities of goods which fail the tests – we can gain some intuition by looking at the extent to which passing these tests varies and changes elasticities across different groups of products.

Table 4 disaggregates the impact of passing the tests on the estimates by 1-digit SITC code. The idea is that any unmodeled tendency to fail the tests is likely to be shared by broad industry

[31] Broda and Weinstein (2006) discuss this issue, and even set the precedent of truncating elasticity estimates to 100.

[32] Surprisingly, the sign of these biases is opposite those from our Monte Carlo Analysis. In settings with multiple endogenous regressors and multiple instruments, how the failure of the exclusion restriction translates into bias is a complicated function of the data structure, and the errors for good 6997 which formed the basis of our Monte Carlo Analyses may be different than the "typical" structure in the data. It is also possible that some of the difference in elasticities is due to compositional differences of goods which pass vs fail the tests.

groups, so by disaggregating in this way we can gain some insight about the role of bias detected by the tests versus unobserved differences shared by broader industry groups. In the first column we present the SITC 1-digit codes. We subdivide estimates for each code depending on whether they pass or fail our tests, in the second row we present the number of goods in each group, and in the third and fourth columns we describe the share of a given group that pass each of the tests. The second row for every product group is goods that pass both tests, so that the share which pass each test is always 100%. In the last 5 columns, we present the median test statistics within each group and the median supply and demand elasticities, as well as $\hat{\rho}$.

Table 4: Impact of the Weak Instruments and Sargan Tests Disaggregated by SITC1 Code

		Pass Tests		Median				
SITC1	Products	J-test	F-test	$J^{\text{P-value}}$	F^{KP}	$\hat{\sigma}$	$\hat{\omega}$	$\hat{\rho}$
0	48	19%	71%	0.974	3.39	2.824	0.189	0.146
0	29	100%	100%	0.193	13.40	1.712	0.098	0.071
1	6	0%	100%	1.000	10.83	2.790	0.001	0.002
1	5	100%	100%	0.477	42.63	5.056	0.556	0.504
2	46	50%	41%	0.878	1.81	2.516	0.001	0.007
2	27	100%	100%	0.148	11.33	2.595	0.605	0.237
3	5	40%	40%	0.960	1.21	9.999	0.791	0.784
3	10	100%	100%	0.087	6.81	3.555	0.067	0.094
4	8	50%	50%	0.847	3.83	5.078	0.143	0.059
4	2	100%	100%	0.024	33.30	1.492	50.095	0.214
5	57	11%	74%	0.996	5.84	3.267	0.045	0.014
5	32	100%	100%	0.272	16.31	1.796	0.546	0.170
6	105	25%	70%	0.979	7.16	2.792	0.025	0.017
6	70	100%	100%	0.093	11.20	1.806	0.195	0.065
7	56	32%	54%	0.906	3.05	2.639	0.000	0.001
7	71	100%	100%	0.105	26.88	2.850	0.344	0.146
8	35	29%	57%	0.926	3.18	2.864	0.114	0.038
8	41	100%	100%	0.149	35.09	2.699	0.242	0.213
9	3	33%	67%	0.947	2.28	4.625	0.646	0.474

Notes: F^{KP} is the median Kleibergen-Paap F-statistic, $J^{\text{P-value}}$ is the p-value of the J-statistic. Products are the count of 4-digit SITC industries. Median estimates across designated products for demand and supply are denoted by $\hat{\sigma}$, $\hat{\omega}$, and $\hat{\rho}$. For the sake of space, robust standard errors from HLIML are available with published estimates. F^{KP} is the median Kleibergen-Paap F-statistic, $J^{\text{P-value}}$ is the p-value of the J-statistic. Pass versus Fail follow our rules of thumb described in Table 2.

In examining the results in Table 4, we need to be careful about the sample size – some 1-digit SITC industries have comparatively few component 4-digit SITC industries. This makes it hard to distinguish correlations between estimated parameters and passing the tests vs small sample issues.

For that reason, we will focus attention on the SITC 1-digit industry codes "0", "2", "5", "6", "7", and "8".

There is substantial variation in the rate at which estimates for SITC 4-digit products fail the tests across different 1-digit industries. This suggests that there are differences in the economic fundamentals driving shocks across different categories of goods, which make them more or less suitable to the method, and that the tests pick these up. However, the broad patterns in estimated $\hat{\sigma}$, $\hat{\omega}$, and $\hat{\rho}$ for estimates that pass versus those that do not are generally similar across 4-digit products within all 1-digit industries as they are in the overall data. Industry "2" is a weak exception to the pattern in $\hat{\sigma}$; industry "0" is an exception to the pattern in $\hat{\omega}$ and $\hat{\rho}$. Furthermore, the magnitude of the effects of passing versus failing the tests are similar across 1-digit industries as in our overall data. This suggests that at least part of the change in overall estimates reflects elimination of bias and not just cross-sectional differences in products.

Table 5 presents a few examples of specific goods that we find helpful for diagnosis. Consider first SITC 7924, which has the description "Aircraft exceeding 15000KG". This good arguably performs the worst of any good in our data set. The F-statistic is 1.18 and the p-value of the J-statistic is 0.94. Neither are close to our rules of thumb. These are potentially not surprising numbers given the product in question. Only five developed countries sparsely export this good to the US. Additionally, violations of the underlying assumptions through correlated supply and demand shocks (e.g., endogenous quality) and common correlated shocks across exporters make a high J-statistic and a low F-statistic unsurprising. These violations seem to push the estimator to the boundary for both supply and demand elasticities returning $\hat{\omega} = 0.001$ and $\hat{\sigma} = 1.214$.

Table 5: Product Level Comparisons

		Passing Test			Elasticity Estimates		
Description	SITC4	$J^{\text{P-val}}$	F^{KP}	Both	$\hat{\sigma}$	$\hat{\omega}$	$\hat{\rho}$
Wood Glue	5837	✗	✓	✗	3.214	0.356	0.368
					(0.025)	(0.025)	(0.031)
Bike Tires	6254	✓	✓	✓	2.099	0.202	0.156
					(0.006)	(0.006)	(0.001)
Articles of Copper, Tin, etc.	6997	✓	✓	✓	2.450	0.484	0.321
					(0.061)	(0.061)	(0.006)
Large Aircraft	7924	✗	✗	✗	1.214	0.001	0.000
					(0.014)	(0.014)	(0.096)

Notes: F^{KP} indicates whether the Kleibergen-Paap F-statistic is above the cutoff rule, $J^{\text{P-val}}$ indicates whether the p-value of the J-statistic is below the cutoff rule. Robust standard errors from HLIML are in parentheses for our estimates of σ, ω, and ρ.

Conversely, SITC 6254, which has the description "New Rubber Tires for Motorcycles and Bicycles" is a well performing product in terms of our metrics. The estimated F-statistic is 2.22

and a J-statistic p-value of 0.06, both passing the statistical tests. The relative homogeneity of this product and dense set of 25 exporters is in line with passing our statistical tests. The resulting supply and demand elasticities are precisely estimated as $\hat{\omega} = 0.202$ and $\hat{\sigma} = 2.099$. The good which we used modeled our Monte Carlo, SITC 6997 "Articles of Copper, Tin, etc," is another well performing good in terms of the statistical tests. The resulting elasticity estimates are also sensible and precisely estimated.

Finally, returning to our pet case – US imports of wood glue SITC 5837. The estimated supply and demand elasticities are $\hat{\omega} = 0.356$ and $\hat{\sigma} = 3.214$, respectively. Visually, Figure 2 suggests sufficient heteroskedasticity for estimation. The F-statistic of the first stage regression for wood glue confirms this assertion. We estimate an F-statistic of 426, which is well above the cutoff. However, the J-statistic returns a value of 38.96, which translates into a 99% probability of rejecting the uncorrelated error hypothesis. Given the marked failure of the J-test, a deeper examination of the underlying model may be warranted.

Table 6: Product Type Comparisons

		Sample Shares		Pass Tests		Elasticity Estimates		
Type	Products	Full	Conditional	$J^{\text{P-val}}$	F^{KP}	$\hat{\sigma}$	$\hat{\omega}$	$\hat{\rho}$
Full Sample	369	56%	56%	27%	63%	3.094	0.030	0.014
						(0.232)	(0.232)	(0.034)
	287	44%	44%	100%	100%	2.206	0.243	0.121
						(0.054)	(0.054)	(0.012)
Differentiated Goods	172	26%	61%	27%	63%	3.162	0.053	0.026
						(0.301)	(0.945)	(0.038)
	109	17%	39%	100%	100%	1.896	0.152	0.105
						(0.029)	(0.069)	(0.010)
Homogeneous Goods	197	30%	53%	27%	63%	2.864	0.001	0.008
						(0.167)	(0.667)	(0.034)
	178	27%	47%	100%	100%	2.441	0.321	0.151
						(0.079)	(0.146)	(0.015)

Notes: Products are the count of 4-digit SITC industries. Sample Shares are calculated across the full sample of 656 products for Full, and within the product type for Conditional. F^{KP} indicates whether the Kleibergen-Paap F-statistic is above the cutoff rule, $J^{\text{P-val}}$ indicates whether the p-value of the J-statistic is below the cutoff rule. The median robust standard errors from HLIML are in parentheses for our estimates of σ, ω, and ρ.

In order to make cross-good comparisons in a more systematic way, we next turn to the Rauch Classifications (Rauch (1999)). We subdivide goods into those that are traded on exchanges (which we refer to as "Homogeneous Goods") and those that are reference priced and differentiated under the Rauch Classifications (which we refer to as "Differentiated Goods").[33] Table 6 shows that homogeneous goods are more likely to pass the tests than differentiated ones: 47% of the homogeneous

[33]We combine reference-priced and differentiated goods into a single group to improve the sample size, as relatively few goods are classified as "differentiated" in the Rauch Classifications.

goods pass both tests versus 39% of differentiated goods. This suggests that there is likely some selection at work in the differences in elasticities across goods that pass versus those that do not. This is unsurprising if we believe that more customized or differentiated goods are more likely to exhibit correlations between variety-specific demand and supply shocks (e.g., endogenous quality upgrading) than more standardized goods traded on exchanges.

Not only do the patterns of which goods pass and fail the tests line up with the Rauch Classification, but Table 6 shows some interesting patterns in goods that pass vs fail the tests even within Rauch Classifications. First, estimates of σ are consistently lower and estimates of ω are consistently higher for goods which pass the tests, even conditional on the Rauch Classification. This suggests reduction of bias, and not just selection in the types of goods that pass. Second, we have the intuition that σ should be lower for homogeneous vs differentiated goods. This pattern is true for goods that pass the test, but not for those which do not – this also strongly suggests that at least some of the difference in estimated elasticities between goods which pass and those which do not is attributable to bias.

5.3 Pass and Fail: Next Steps

Thus far, we have established that the tests help us identify a set of goods for which elasticity estimates following the Feenstra (1994) method are credible. In this section, we now turn our attention to goods that fail the statistical tests – what can we say about these goods, and how should a researcher go about obtaining elasticity estimates?

For goods that fail the F-test, one approach for inference would be to compute weak instrument robust standard errors. Even the robust HLIML standard errors we present are likely to underestimate the uncertainty in point estimates in the presence of weak instruments. Computing weak instrument robust standard errors will more accurately convey the uncertainty in these estimates. One advantage of this approach is that it would allow the use of estimates (and product categories) for which the instruments are weak. The main downside is that it is computationally costly, which is an important issue when estimating elasticities for thousands of goods. And secondarily, the researcher also needs to accurately account for this uncertainty in any application of the estimates. There is not a good tradition of taking into account the standard errors for estimated elasticities in the trade literature – for instance, in expressing how that uncertainty passes through to later estimation stages that depend on elasticities, e.g., in gains from trade – and so information in this form would be more difficult for the field to incorporate.

An alternative approach which might help in the case of failing either the F- or the J- tests would be to consider sources of model misspecification. If the Feenstra (1994) model is misspecified, this could cause failures of the J-test, if, for instance, there is an unmodelled fundamental linking supply and demand that causes correlation of the supply and demand residuals. Misspecification could also make the measurement of the hyperbolae in the first stage weak and lower the first-stage F-statistics. Models that resolve the misspecification by capturing the missing mechanisms structurally may help solve these problems. However, this is not necessarily so: additional controls in extensions of Feenstra (1994) may soak up some of the heteroskedasticity and yield hyperbolae which are harder to differentiate. In this case more structure may exacerbate the weak instruments problem. And more generally, if the problem is simply that the structure of supply and demand shocks is too similar across supplying countries, greater structure is unlikely to resolve the problem and researchers may have to consider approaches beyond heteroskedastic estimation such as more conventional IV approaches.

While there is no guarantee that more structure will resolve failures of the tests, it is one natural route to consider. There are many extensions of the Feenstra (1994) to encompass other potential forces. We direct the reader particularly toward Feenstra and Romalis (2014) (endogenous quality), Feenstra and Weinstein (2017) (endogenous markups with translog preferences), Feenstra et al. (2018) (nested demand systems) and Farrokhi and Soderbery (2022) (trade elasticities in general equilibrium). All of these papers develop heteroskedastic estimators taking into account more complicated economic structures, which may have stronger first stages or better control for omitted variables that could translate in the Feenstra (1994) setting as correlation in the supply and demand shocks.

One extension of the Feenstra (1994) framework that has received considerable attention in the literature is endogenous quality as in Feenstra and Romalis (2014), and we thus focus on this extension. We show in the Appendix that the model of Feenstra (1994) is equivalent to a time-differenced special case of the model of Feenstra and Romalis (2014) in which firms cannot endogenously choose their quality and the measure of exporters is exogenous. Therefore, if Feenstra and Romalis (2014) is the true model, then our estimates applying Feenstra (1994) should be biased due to the omitted variables driving the choice of quality. Goods that pass the statistical tests, and in particular the J-test, should be those for which the endogenous quality channel is weak, and in this case the true supply and demand elasticities from the Feenstra and Romalis (2014) framework will be close to those that we estimate using the Feenstra (1994) model.

Leaving the derivations for the Appendix, we observe that Feenstra and Romalis (2014)'s estimating equation can be written as:

$$(\Delta^k p_{it})^2 = \frac{\omega_{FR}*\theta_{FR}*(\sigma_{FR}-1)}{(1+\omega_{FR})(\sigma_{FR}-1)*(1+\gamma_{FR})}\left(\Delta^k s_{it}\right)^2 + \frac{\omega_{FR}(\sigma_{FR}-2)-1}{(1+\omega_{FR})(\sigma_{FR}-1)}\Delta^k s_{it}\Delta^k p_{it} + f(\tau_{it}, p^*_{it}, F_i, L_i) + u^{FR}_{it}, \tag{9}$$

where the parameter θ_{FR} governs quality adjustments, γ_{FR} is the Pareto shape parameter governing firm productivity, and $f(\tau_{it}, p^*_{it}, F_i, L_i)$ is a complicated function of trade costs (τ), shipped prices (p^*), fixed costs (F), and employment (L) that Feenstra and Romalis (2014) leverage to identify supply, demand and quality parameters. As firms become homogenous along productivity and quality (i.e., $\theta_{FR} \to 0$ and $\gamma_{FR} \to 0$), we show in the Appendix that $f(\tau_{it}, p^*_{it}, F_i, L_i) \to 0$ and the estimating Equation (9) is identical to (5) as the coefficients in front of prices and shares become identical.

If Feenstra and Romalis (2014) is the true model, we can see how our estimates applying Feenstra (1994) should be biased due to omitted variables (i.e., $f(\tau_{it}, p^*_{it}, F_i, L_i)$) and endogeneity implied by model misspecification (i.e., through θ_{FR} and γ_{FR}). Our assertion is that passing the statistical tests, specifically the J-test, should align Feenstra (1994) with Feenstra and Romalis (2014) estimates with the same underlying data.

Direct parameter comparisons of σ and ω across the two models are misleading because these parameters capture different things in the two models. In Feenstra and Romalis (2014) demand and supply elasticities are convoluted by firm heterogeneity and quality. However, the parameter governing vertical shifts in supply represents the same fundamental mechanism across both models. We show Feenstra and Romalis (2014) implies the parameter $\rho_{FR} \equiv \frac{\omega_{FR}(\sigma_{FR}-1)}{1+\omega_{FR}\left(1+\frac{(\sigma_{FR}-1)(1+\gamma_{FR})}{1+\theta_{FR}(\sigma_{FR}-1)}\right)}$ governs shifts in their supply curves, and is directly comparable to Feenstra (1994)'s $\rho \equiv \frac{\omega(\sigma-1)}{1+\omega\sigma}$. In order to make comparisons across supply and demand elasticity estimates we define the productivity distribution adjusted supply elasticity $S_{FR} \equiv \frac{\omega_{FR}}{1+\omega_{FR}}(1+\gamma_{FR})$ which we compare to ω and the quality adjusted demand elasticity $D_{FR} \equiv \frac{\sigma_{FR}-1}{1+\theta_{FR}(\sigma_{FR}-1)}$ which we compare to ω.

Table 7 compares our estimates applying Feenstra (1994) to Feenstra and Romalis (2014) estimates both conditional on passing and failing the J and F tests. We present results for the full sample and then disaggregated by the Rauch Classification. In the first four columns we summarize the relevant sample, the number of products that pass and fail the tests within the given sample, and the percentage of each subset that pass the J- and F-tests. In the remaining columns we compare parameter estimates from Feenstra (1994) to the analogue from Feenstra and Romalis (2014): columns five and six compare ρ, columns seven and eight compare supply elasticities, and columns nine and ten compare demand elasticities. When comparing parameters, columns five, seven, and nine present the median value from Feenstra and Romalis (2014), while columns six, eight, and ten present the percentage difference in the estimates between Feenstra (1994) and Feenstra and Romalis (2014).

Table 7: Relation of Tests Disaggregated by SITC1 to Feenstra and Romalis (2014) Estimates

				Feenstra and Romalis (2014)					
		Pass Tests		Supply Shifter		Supply Elasticity		Demand Elasticity	
Sample	Products	J-test	F-test	$\hat{\rho}_{FR}$	$\frac{\hat{\rho}-\hat{\rho}_{FR}}{1/2(\hat{\rho}_{FR}+\hat{\rho})}$	$\hat{S}_{FR}$	$\frac{\frac{\hat{\omega}}{1+\hat{\omega}}-\hat{S}_{FR}}{1/2\left(\hat{S}_{FR}+\frac{\hat{\omega}}{1+\hat{\omega}}\right)}$	$\hat{D}_{FR}$	$\frac{\hat{\sigma}-\hat{D}_{FR}}{1/2(\hat{D}_{FR}+\hat{\sigma})}$
All goods	369	27%	63%	0.194	−0.501	0.778	−0.972	2.222	0.330
	287	100%	100%	0.197	−0.121	0.761	−0.706	2.233	0.068
Differentiated	172	27%	63%	0.170	−0.286	0.637	−0.768	2.221	0.300
	109	100%	100%	0.148	−0.047	0.483	−0.532	2.181	−0.020
Homogeneous	197	27%	63%	0.214	−0.688	0.933	−1.151	2.227	0.355
	178	100%	100%	0.227	−0.167	1.060	−0.813	2.253	0.121

Notes: Median estimates of $\rho_{FR} \equiv \frac{\omega_{FR}(\sigma_{FR}-1)}{1+\omega_{FR}\left(1+\frac{(\sigma_{FR}-1)(1+\gamma_{FR})}{1+\theta_{FR}(\sigma_{FR}-1)}\right)}$, $S_{FR} \equiv \frac{\omega_{FR}}{1+\omega_{FR}}(1+\gamma)$, and $D_{FR} \equiv \frac{\sigma_{FR}-1}{1+\theta_{FR}*(\sigma_{FR}-1)}$ are presented in Columns 5,7,9. Percentage differences between Feenstra and Romalis (2014) and our estimates are presented in Columns 6,8,10. "Differentiated" indicates differentiated and reference priced goods and "Homogeneous" indicates homogeneous goods following Rauch (1999)'s conservative classification.

Table 7 shows that for goods passing the tests, estimates using the Feenstra (1994) method are much closer to those from the Feenstra and Romalis (2014) method relative to estimates for goods which fail the tests. Over the full sample, we see the difference in the median ρ falls from 50% to 12%, the difference in the median demand elasticity falls from 33% to 7%, while the difference in the median supply elasticity falls from 97% to 71%. We observe similar results when we disaggregate by Rauch (1999) Classification. Notably, estimates for Differentiated goods passing the tests are closer to the estimates from Feenstra and Romalis (2014) when compared to estimates for Homogeneous goods. This is consistent with the idea that endogenous quality is a particularly important force for differentiated goods.[34]

Table 7 suggests that endogenous quality is potentially an important source of misspecification of the Feenstra (1994) model – when the estimates pass the tests they are much closer to those of Feenstra and Romalis (2014) than when they do not. However, it also suggests that endogenous quality is not the only potential source of misspecification, and furthermore that estimates from Feenstra and Romalis (2014) may also be affected by some of the biases that we identify. Even conditional on passing the tests, the estimates do not line up fully with Feenstra and Romalis (2014). This is not surprising: endogenous quality is not the only reason that the model could be misspecified. And

[34]In the Appendix, we present our results disaggregated by SITC 1-digit code. We observe similar patterns in the differences for all three elasticities: in general, they are better correlated with Feenstra and Romalis (2014) conditional on passing the tests, but this is not universally true.

more generally, even if Feenstra and Romalis (2014) helps resolve possible violations of the exclusion restriction, it too may suffer from weak instrument problems. Thus, simply switching to estimates from Feenstra and Romalis (2014) is unlikely to deal with all of the issues identified by the tests. The good news is that we can, and should, apply exactly the same tests to estimates arising from Feenstra and Romalis (2014), or any other extension of the basic framework. Even though no single model is likely to work perfectly for every good, we can get a probabilistic measure of which models are doing a good job of solving exclusion restriction violations or weak instruments problems, for which types of goods. Finally, we find it reassuring that the patterns of statistical violations tend to follow our economic intuition, suggesting heteroskedastic estimation in the style of Feenstra (1994) can successfully estimate elasticities with an appropriate structural framework.

6 Conclusion

The Feenstra (1994) method is widely used in the trade literature to estimate elasticities which are critical to answering questions as broad as the gains from trade and the impacts of trade policy. In this paper, we examine the key identification criteria necessary for the method to yield unbiased estimates of the elasticities of interest. We provide more intuition about how the estimator is working and where the estimation is coming from; in particular, that Feenstra (1994) can be understood as IV strategy for Leamer (1981) hyperbolae. Using this interpretation permits us to map the Feenstra (1994) identification criteria to the usual requirements for an unbiased IV strategy: the exclusion restriction must be satisfied and there must be a strong first stage.

We show through Monte Carlo analysis that this understanding is the right one: that violating the condition analogous to the exclusion restriction causes the estimates to fail a Sargan J-test, while violating the condition analogous to a strong first-stage causes the estimates to fail a first-stage F-test. Furthermore, we show that if we restrict attention to the subset of estimates which pass both tests, there is less median bias in these results than in the other estimates in our Monte Carlo analysis.

Finally, we turn to U.S. import data over the years 1984-2011 at the SITC 4-digit level. We show that for these data, elasticity estimates for a substantial portion of goods pass the tests, but also estimates for many goods do not pass. Patterns in the resulting estimates of supply and demand elasticities (i.e., estimates passing both statistical tests exhibit larger export supply elasticities and lower import demand elasticities) are consistent across SITC 1-digit industries. We also subdivide goods by their Rauch Classifications. In doing so, we observe a similar pattern in our elasticity estimates for goods which pass the tests versus those which do not, which is further evidence that the tests are eliminating bias. However, differentiated goods are less likely to pass the tests than homogeneous goods, and both tend to have different elasticities conditional on passing the tests. This suggests that selection explains some of the difference in elasticities between goods which pass the tests and those which do not.

We hope that implementing these tests will become a matter of standard practice for practitioners using these methods and elasticities. More broadly, despite its utility, the field has made less use of the Feenstra (1994) method than it could. We see this as a product both because the intuition of the estimator is difficult to grasp and because it is hard to know whether the structural assumptions are satisfied in practice. We hope that this paper will help alleviate both concerns, by providing a more intuitive understanding of how the estimator works and by providing tests of the structural assumptions which can identify when the assumptions have been violated.

References

Andrews, I., Stock, J., and Sun, L. (2019). Testing for Weak Instruments in Linear IV Regression. *Annual Review of Economics.*

Arkolakis, C., Costinot, A., and Rodriguez-Clare, A. (2012). New Trade Models, Same Old Gains? *American Economic Review*, 102(1):94–130.

Armington, P. S. (1969). A theory of demand for products distinguished by place of production. *Staff Papers - International Monetary Fund*, 16(1):159–178.

Broda, C., Limao, N., and Weinstein, D. (2008). Optimal tariffs and market power: the evidence. *The American Economic Review*, 98(5):2032–2065.

Broda, C. and Weinstein, D. (2006). Globalization and the gains from variety. *Quarterly Journal of Economics*, 121(2):541–585.

Eaton, J. and Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5):1741–1779.

Farrokhi, F. and Soderbery, A. (2022). Trade Elasticities in General Equilibrium: Demand, Supply, and Aggregation. *Mimeo.*

Feenstra, R. C. (1994). New product varieties and the measurement of international prices. *The American Economic Review*, 84(1):157–177.

Feenstra, R. C., Luck, P., Obstfeld, M., and Russ, K. N. (2018). In search of the armington elasticity. *Review of Economics and Statistics*, 100(1):135–150.

Feenstra, R. C. and Romalis, J. (2014). International Prices and Endogenous Quality. *Quarterly Journal of Economics*, 2(129):477 – 527.

Feenstra, R. C. and Weinstein, D. E. (2017). Globalization, Markups, and US Welfare. *Journal of Political Economy*, 125(4):1040–1074.

Ferguson, S. M. and Smith, A. (May 2022). Import Demand Elasticities Based on Quantity Data: Theory, Evidence and Implications for the Gains from Trade. *Canadian Journal of Economics.*

Grant, M. (2020). Why Special Economic Zones? Using Trade Policy to Discriminate across Importers. *American Economic Review*, 110(5):1540–1571.

Hausman, J. A., Newey, W. K., Woutersen, T., Chao, J. C., and Swanson, N. R. (2012). Instrumental variable estimation with heteroskedasticity and many instruments. *Quantitative Economics*, 3(2):211–255.

Hillberry, R. and Hummels, D. (2013). Trade Elasticity Parameters for a Computable General Equilibrium Model. In Jorgenson, D. and Dixon, P., editors, *Handbook of Computable General Equilibrium Modelling*, pages 1213–1269. Elsevier.

Hottman, C. J., Redding, S. J., and Weinstein, D. E. (2016). Quantifying the Sources of Firm Heterogeneity. *The Quarterly Journal of Economics*, 131(3):1291 – 1364.

Leamer, E. (1981). Is it a demand curve, or is it a supply curve? Partial identification through inequality constraints. *The Review of Economics and Statistics*, 63(3):319–327.

Nelson, C. R. and Startz, R. (1990). Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator. *Econometrica*, 58(4):967.

Ossa, R. (2014). Trade Wars and Trade Talks with Data. *American Economic Review*, 104(12):4104 – 4146.

Parente, P. and Silva, S. (2012). A cautionary note on tests of overidentifying restrictions. *Economics Letters*, 115(2):314–317.

Rauch, J. (1999). Neworks versus markets in international trade. *Jounal of International Economics*, 48(1):7–37.

Rigobon, R. (2003). Identification through heteroskedasticity. *Review of Economics and Statistics*, 85(4):777 – 792.

Sargan, J. D. (1958). The Estimation of Economic Relationships using Instrumental Variables. *Econometrica*, 26(3):393.

Soderbery, A. (2010). Investigating the asymptotic properties of import elasticity estimates. *Economics Letters*, 109:57–62.

Soderbery, A. (2015). Estimating import supply and demand elasticities: Analysis and implications. *Journal of International Economics*, 96(1):1–17.

Soderbery, A. (2018). Trade elasticities, heterogeneity, and optimal tariffs. *Journal of International Economics*, 114:44–62.

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586.

Stock, J. H. and Yogo, M. (2005). Testing for weak instruments in linear iv regression. In Andrews, D. W. K. and Stock, J. H., editors, *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, chapter 5. Cambridge University Press.

Appendix

I Economic and Econometric Intuition of Leamer (1981)

To gain greater intuition for the Feenstra (1994) method, we turn to its theoretical foundation: Leamer (1981). Leamer (1981) assumes a reduced-form constant elasticity of supply and demand,

$$\begin{aligned} Q_{it} &= \alpha_i + \beta P_{it} + \epsilon_{it} && \text{(Demand)} \\ Q_{it} &= \varphi_i + \gamma P_{it} + \nu_{it} && \text{(Supply).} \end{aligned} \tag{10}$$

where Q_{it} and P_{it} are quantity and prices, respectively for good i in period t, and the elasticities of supply and demand are γ and β, respectively. Although the Leamer (1981) method cannot point identify supply and demand elasticities, it does set identify them. Under the assumption that supply and demand shocks (ν_{it} and ϵ_{it}) are uncorrelated, it is possible to constrain supply and demand elasticities to a hyperbola.

Equilibrium price and quantity are functions of the supply and demand shocks and the supply and demand elasticities. Thus, the variances in price and quantity are functions of the elasticities and the variances in the supply and demand shocks. Furthermore, there will be a non-zero covariance between price and quantity even though the supply and demand shocks are uncorrelated, as each of price and quantity is a function of both supply and demand shocks, and the magnitude of the covariance will reflect the underlying variances and the respective elasticities. This creates a system of three equations (two for the variances and one for the covariance) as a function of four unknowns (the variances of both shocks and the elasticities of supply and demand). Leamer (1981) shows this set of equations can be solved to constrain the elasticities of supply and demand to fall on a hyperbola defined by the variances of price and quantity and their covariance.

Let $\mathcal{s}_{pq}$ denote the sample covariance of price and quantity, $\mathcal{s}_{pp}$ the sample variance of price, and $\mathcal{s}_{qq}$ the sample variance of quantity. The equation for this hyperbola is related to the coefficient of a regression of price on quantity and the coefficient of the "reverse" regression. Define the coefficient of a regression of price on quantity as $b \equiv \frac{\mathcal{s}_{pq}}{\mathcal{s}_{pp}}$, the coefficient of the reverse regression as $b_r \equiv \frac{\mathcal{s}_{qq}}{\mathcal{s}_{pq}}$, and their sample R-squared statistic $R^2 \equiv \frac{\mathcal{s}_{pq}^2}{\mathcal{s}_{pp}\mathcal{s}_{qq}}$. Then the Leamer (1981) hyperbola can be expressed as

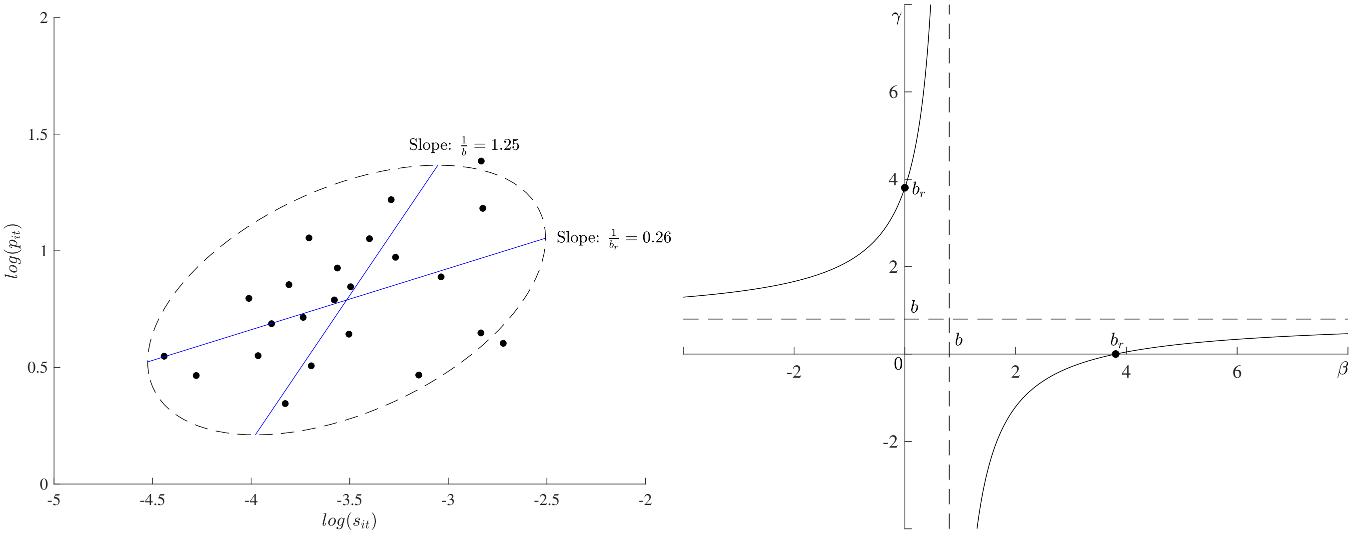
$$(\hat{\beta} - b)(\hat{\gamma} - b) = (R^2 - 1)(b * b_r).$$

We show how this process works in practice by turning to the price and quantity of U.S. imports of the 4-digit SITC code 5837 from the UK between 1984-2011 (these are various polyvinyl acetate wood glues).[35] For UK wood glue we estimate $b = 0.8$, $b_r = 3.8$ and $R^2 = 0.21$. Figure 3 presents the analysis. Panel (a) is the cloud of raw data with fitted lines corresponding to the direct and reverse regressions and an ellipse encasing the data.[36] Panel (b) presents the resulting hyperbola. The key point of Leamer (1981) is that given the variation of prices in quantities in the cloud of data, the hyperbola contains the true pair of elasticity estimates. To see this clearly, if we were to suppose the slope of demand was known to be $\hat{\beta} = 0$ then we would also know the slope of the supply curve is $\hat{\gamma} = b_r = 3.8$. Furthermore, assuming the signs of the supply and demand elasticities restricts the range of possible elasticities further. Figure 4a presents this subset (following Leamer (1981) we impose a positive slope for supply and a negative slope for demand). From the truncated hyperbola we can tell that while any demand elasticity such that $\hat{\beta} \leq 0$ is possible given the data, the true estimate of the supply elasticity must satisfy $\hat{\gamma} \in [b, b_r]$. For UK wood glue in particular, we have $\hat{\beta} \leq 0$ and $\hat{\gamma} \in [0.8, 3.8]$ as shown in Figure 4a.

[35] The dataset we use for these and other data in this paper is introduced in Section 4 of the paper.

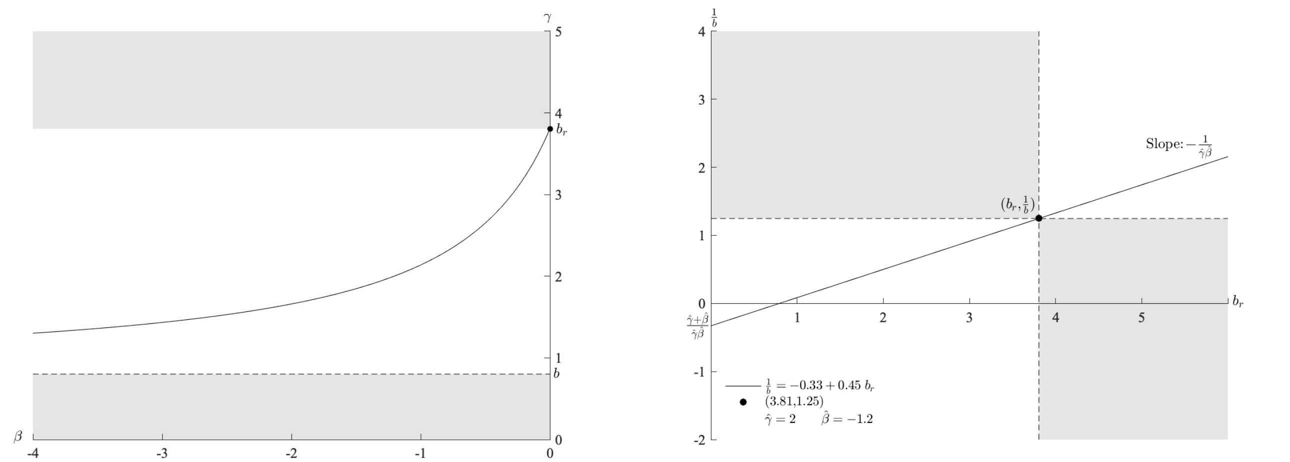
[36] The ellipse is the iso-contour of the normal distribution and represents the 90% confidence interval of the data. Informally, it is related to the R^2 of the simple regressions, and embodies the central tendency of the data. Formally, the ellipse is constructed by taking the eigen decomposition of the covariance matrix, where the eigenvectors capture the direction of greatest variance, and the eigenvalues represent the scale of the variance.

Figure 3: UK Exports of SITC 5837



(a) Cloud of data (b) Hyperbola

Figure 4: UK Exports of SITC 5837



(a) Truncated hyperbola (b) Hyperbola as data

Additionally, we can rearrange the Leamer (1981) hyperbola to represent it as a point in $(b_r, \frac{1}{b})$ space as,

$$\frac{1}{b} = \frac{\hat{\gamma} + \hat{\beta}}{\hat{\gamma}\hat{\beta}} - \frac{1}{\hat{\gamma}\hat{\beta}} b_r. \tag{11}$$

Here, rather than the true elasticities represented as a point on a hyperbola, they must combine to form a line that passes through the point $(b_r, \frac{1}{b})$. Figure 4b presents this point for UK exports of wood glue along with a hypothetical pair of elasticity estimates ($\hat{\beta} = -1.2$ and $\hat{\gamma} = 2$) that generate a line satisfying the condition. The infeasible region (i.e., where a line passing through the point is impossible unless either supply slopes down or demand slopes up) has also been grayed out.

II Testing of Heteroskedasticity

In the text, we employ tests to detect weak instruments, which is strongly recommended because weak instruments are a well-known source of problems in econometric inference. In this section, we outline how to test for the simple presence of heteroskedasticity, which is equivalent to the existence of a first stage. First, we provide and empirical test on the data which yields a probability that the necessary condition is satisfied for a given product category.

As we make clear in the text, the existence of a first stage is simply a requirement of statistically different hyperbolae in the data. Thus, as a practical matter, rather than testing the relative distributions of the variances in the unobservable shocks, we can establish an equivalent necessary condition by testing whether the hyperbolae are different directly. Of course, this raises the problem of finite samples: due to inherent noise in the data, two hyperbolae which are in fact the same could have different sample variances and covariances. In this case, the elasticities would be identified purely off of noise, and would be inconsistent.[37] Since both b and b_r will be different if the hyperbola are different, it is possible to test any combination to establish differences in the hyperbolae.

The test we wish to run is similar in spirit to an F-test. We wish to test the hypothesis that $b_i \cdot b_{i,r}$ are distinct across source countries i relative to the null hypothesis that $b_i \cdot b_{i,r}$ is shared for all countries i. In fact, if the sample analogue $\hat{b}_i \cdot \hat{b}_{i,r}$ were distributed chi-squared and these chi-squared distributions were independent across sources i, then the distribution would be F-distributed and we could perform a standard F-test. However, it is impossible that this ratio would be distributed this way. Notice, $\Delta^k S_{it}$ and $\Delta^k P_{it}$ are correlated since they depend on the same supply and demand elasticities (with the covariance driven by the supply and demand elasticities).[38] Depending on the assumption on correlations of shocks across countries, there may be other objections as well.

Given the complicated correlations in the data (which are both unknown and a function of unknown parameters), we bootstrap the test of heteroskedasticity to account for the distributions of $\Delta^k S_{it}$ and $\Delta^k P_{it}$ in the data, including potential correlation patterns across countries. We perform a block bootstrap at the good-year level with replacement. This means for each repetition for each good we draw 29 years (to match the 29 years in our sample after accounting for differences) with replacement. When we draw a year, we adopt all the suppliers of that good in a given year. We may draw the same year in the data multiple times, while other years in the data may not be drawn at all for a given repetition. Given our simulated data, we then calculate $\hat{b}_i \cdot \hat{b}_{i,r}$ for each supplier i as well as $\hat{b} \cdot \hat{b}_r$ pooling data from all suppliers. We then calculate for the repetition w the statistic Φ_w according to

$$\Phi_w = \frac{\sum_i n_{iw} \cdot \hat{b}_i \cdot \hat{b}_{i,r}}{n_w \cdot \hat{b} \cdot \hat{b}_r},$$

where n_{iw} is the number of observations for source i in repetition w and $n_w = \sum_i n_{iw}$ is the total number of observations.[39] If the value of statistic is $\Phi_w = 1$ for a given repetition, that would imply the ratio of sample variances are the same across all supplying countries, which would be more likely under homoskedasticity than heteroskedasticity. We repeat this procedure 5,000 times, yielding an empirical distribution of the test statistic. Using this distribution, we then test the alternative hypothesis that $\Phi \neq 1$ against the null hypothesis that $\Phi = 1$ using a two-tailed test.

In practice, if the median value of the statistic across all repetitions is greater than 1, we take the share of draws less than one and divide by half (as we conduct a two-tailed test) – this is the probability level at which we reject the null hypothesis. We do the same thing if the median of the statistic across all repetitions is less than 1, except we find the share of repetitions for which the statistic is greater than 1. We also find the mean of the statistic and normalize it

[37] In fact, this was the Frisch (1933) critique of a similar method in Leontief (1929). In Leontief (1929), the sample was simply divided in two without any rationale for the necessary heteroskedasticity; this is in contrast to Feenstra (1994) where heteroskedasticity is more pluasible due to differences in supplier.

[38] For $\hat{b}_i \cdot \hat{b}_{i,r}$ to be distributed chi-squared, we would have to assume $\Delta^k a_{jt}$ and $\Delta^k B_{jt}$ were correlated, contrary to the assumptions underlying the Feenstra (1994) model.

[39] Note that n_{iw} need not be 29, as in some years some suppliers may not export, and n_{iw} will vary from repetition to repetition depending on which years from the data we draw for a given repetition.

by the standard deviation to capture how much heteroskedasticity the bootstrap finds in the data (for both the mean and standard deviations we trim the top and bottom 2.5% of realizations; this yields a clearer picture as sometimes extreme values yield very large standard deviations, although this is not necessary).

We find that the data are generally consistent with heteroskedasticity as assumed in Feenstra (1994). The probability of the null for the median product is 0.21%, and the average across all products is 2.6%. For more than 90% of products, the probability of the null is under 10%.

III Information about Monte Carlo Analysis

In this section of the Appendix, we provide more information about the construction of the Monte Carlo Analysis. In the first subsection we provide more information about SITC 6997 and how it compares to other goods in out dataset. And in the second subsection we provide more information about how we draw supply and demand residuals for the Monte Carlos.

III.i Information about SITC 6997

SITC 6997 is used to generate the moments underlying our Monte Carlo analyses. In this part of the Appendix, we compare some moments of SITC 6997 to the distribution of moments across goods in the dataset.

Table 8: Dataset Moments

		Full Sample Distribution				
Moment	SITC6997	Min	25th	Median	75th	Max
$\hat{\sigma}$	2.45	1	1.54	2.72	6.98	$1.73*10^{45}$
$\hat{\omega}$	0.48	0.0001	0.0001	0.13	1.7	$6.60*10^{44}$
Suppliers	73	4	27	42	56	120
J-test P-value	0.24	0	0.06	0.51	0.98	1
F-stat Cutoff	7.4	0	1.18	4.05	26.62	$5.85*10^{5}$

III.ii Construction of the Monte Carlos

In constructing the Monte Carlo dataset, we wish to replicate some essential features of SITC 6997, and this drives our choice of simulation procedure. We will refer to each draw in the Monte Carlo Analysis as a simulation.

First, we preserve the number of suppliers from the data. This number of suppliers is held fixed in all of our simulations.

Second, we pick a subset of years for each supplier to export and be in the data. We wish for this to match the distribution of years in the data. To do this, we take the mean and standard deviation (across all suppliers in the data) of the number of years for that supplier. Then, in each simulation and for each supplier we draw a number of years from that distribution.

Third, we wish to match the degree of heteroskedasticity in the data. To do this, for each supplier in the data we take the variance of the supply and demand residuals. We then take the mean and variance of these distributions and find Pareto distributions which match these moments. For each supplier and each simulation, we draw from both of these distributions to obtain the variance in suppler and demand residuals.

Fourth, we wish to match the correlation coefficient between the supply and demand residuals. We measure the average correlation between the estimated supply and demand residuals. Then, within each simulation, we randomize the average correlation coefficient between the shocks according to a uniform distribution with a minimum of zero and a maximum of double the average correlation in our data (so that the mean correlation coefficient across all simulations is in expectation equal to the correlation coefficient targeted by the Monte Carlo).

Fifth, we wish to match the distribution of the common shock to all suppliers in a given year in our simulations. The common shock in a given year is the average of the estimated residual across all suppliers of a given good in that year. We find the mean and variance of these shocks.

Sixth, based on the distribution of the common supply and demand shocks and the variance of supply and demand shocks we draw for the given supplier-year pair (following part 3), we find the variance of the idiosyncratic component of the supply and demand shocks so that the combined variance in shocks (idiosyncratic+common components) matches what we intend.

Seventh, we draw common shocks for every simulation-year. We draw the common shock for every year based on the chosen correlation coefficient for the simulation and and mean and standard deviation of the common shocks.

And eighth, we draw the idiosyncratic shocks for every simulation-supplier-year given the means and variances of the idiosyncratic shock for that supplier-simulation and the correlation coefficient for that simulation.

In different Monte Carlos (and described in the main text as MC1-4), we skip some of the above steps and adjust the remaining steps to match the other moments. In MC1, we set the correlation coefficient to zero in all supplier-simulations and the variance of the common supply and demand shocks to zero in all simulations. We then adjust the idiosyncratic shocks so that we still match the overall variance in supply and demand shocks for a given supplier-simulation. In MC2, we do not adjust to correlation coefficient but adjust the variances of the idiosyncratic and common supply and demand shocks for each simulation a described in MC1. For MC3, we make the correlation coefficient 0 for every simulation, but we do not adjust our routine for common and idiosyncratic shocks. And Finally for MC4 we follow steps 1-8 exactly.

IV Nesting of Feenstra (1994) in Feenstra and Romalis (2014)

In this section of the Appendix, we show that the model of Feenstra (1994) is equivalent to a time-differenced special case of the model of Feenstra and Romalis (2014) in which firms cannot endogenously choose their quality and the measure of exporters is exogenous.

Feenstra (1994) shows that the measure of varieties is isomorphic to the taste shifter, so that in order to obtain Feenstra (1994) from Feenstra and Romalis (2014), both the measure of firms and their quality index must be exogenous. In Feenstra and Romalis (2014), the parameter θ captures the elasticity of quality with respect to investments in quality; as θ goes to 0 endogenous quality becomes infinitely costly to produce, and quality is simply an exogenous shock as in Feenstra (1994). And in Feenstra and Romalis (2014), the parameter γ is the shape parameter of the firm productivity distribution. The fixed cost of exporting is a decreasing function of firm productivity, and as γ goes to zero, there is no longer a fixed cost of exporting and the measure of varieties from any given country becomes exogenous as in Feenstra (1994). When both of these conditions are met, the taste shifter in Feenstra and Romalis (2014) becomes exogenous and the model reduces to that of Feenstra (1994).[40]

Under the provided conditions, the estimating equation for Feenstra (1994) follows from taking a time difference of this special case of the Feenstra and Romalis (2014). The supply equation in Feenstra and Romalis (2014) is equation (E5), reproduced below for convenience:

$$
\begin{aligned}
&(1+\omega_2)\left(\ln uv_{it}^{k} - \ln uv_{jt}^{k}\right) - \theta\left(\ln uv_{it}^{*k} - \ln uv_{jt}^{*k}\right) - \omega_2\left(\ln X_{it}^{k} - \ln X_{ij}^{k}\right) \\
&= (\eta_0 - 1)\left(\ln tar_{it}^{k} - \ln tar_{jt}^{k}\right) + \omega_1\left(\ln dist_{i}^{k} - \ln dist_{j}^{k}\right) + \tilde{\xi}_{it}^{k} - \tilde{\xi}_{jt}^{k}
\end{aligned}
$$

[40] Technically, θ must approach 0 faster than γ for the model to be well-behaved as we discuss below.

while the demand equation from FR is (E6) in the online appendices, reproduced below for convenience:

$$\begin{aligned}
&\ln X_{it}^{k} - \ln X_{ij}^{k} + A^{k}\left[\left(\ln uv_{it}^{k} - \ln uv_{jt}^{k}\right) - \alpha^{k}\theta\left(\ln uv_{it}^{*k} - \ln uv_{jt}^{*k}\right)\right] \\
&= \delta_0\left(\ln L_{it} - \ln L_{jt}\right) + \delta_i - \delta_j - B^{k}\beta'\left(F_i^{k} - F_j^{k}\right) - C^{k}\left(\ln tar_{it}^{k} - \ln tar_{jt}^{k}\right) + \tilde{\epsilon}_{it}^{k} - \tilde{\epsilon}_{jt}^{k}
\end{aligned}$$

and in both equations

1. i and j subscripts are source countries, a t subscript denotes the period, a k superscript denotes the destination country, and an $*$ superscript denotes a F.O.B. (as opposed to C.I.F.) price
2. uv denotes the unit value
3. α^k is a taste for quality in the destination k (and is normalized to 1 in the case of the U.S.)
4. θ is the returns to scale in the production of quality, ω_2 is the export supply elasticity (more specifically, it relates specific trade costs to aggregate volume), σ is the elasticity of substitution across varieties, γ is the shape parameter of the Pareto-distributed firm productivities, and $A^k = \dfrac{(\sigma-1)(1+\gamma)}{1+\alpha^k\theta(\sigma-1)}$.
5. tar denotes the (ad-valorem) tariff on imports
6. $\tilde{\xi}$ and $\tilde{\epsilon}$ are errors which are uncorrelated with each other in differences
7. The remaining terms are constants which are defined in FR

We next take the limit as θ,$\gamma \to 0$. However, we must note a minor technical point: in order to obtain a finite quality index, Feenstra and Romalis (2014) assume that $\gamma > \alpha^k\theta(\sigma-1)$. This means that in our limit, θ must approach 0 more quickly than $\dfrac{\gamma}{\alpha^k(\sigma-1)}$. This is readily accomplished by, e.g., setting $\dfrac{\gamma}{\alpha^k(\sigma-1)} = t$ and$\theta = t^2$ for $t < 1$ and taking the limit as $t \to 0$.

With this technical point resolved, under the limit as $\theta, \gamma \to 0$, then $A^k \to (\sigma-1)$ and the supply equation converges to

$$\begin{aligned}
&(1+\omega_2)\left(\ln uv_{it}^{k} - \ln uv_{jt}^{k}\right) - \omega_2\left(\ln X_{it}^{k} - \ln X_{ij}^{k}\right) \\
&= (\eta_0 - 1)\left(\ln tar_{it}^{k} - \ln tar_{jt}^{k}\right) + \omega_1\left(\ln dist_i^{k} - \ln dist_j^{k}\right) + \tilde{\xi}_{it}^{k} - \tilde{\xi}_{jt}^{k}
\end{aligned}$$

while the demand equation converges to (noting that $C^k \to (\sigma-1)$ in this limit)

$$\begin{aligned}
&\ln X_{it}^{k} - \ln X_{ij}^{k} + (\sigma-1)\left(\ln uv_{it}^{k} - \ln uv_{jt}^{k}\right) \\
&= \delta_0\left(\ln L_{it} - \ln L_{jt}\right) + \delta_i - \delta_j - B^{k}\beta'\left(F_i^{k} - F_j^{k}\right) - (\sigma-1)\left(\ln tar_{it}^{k} - \ln tar_{jt}^{k}\right) + \tilde{\epsilon}_{it}^{k} - \tilde{\epsilon}_{jt}^{k}
\end{aligned}$$

and if we take a time difference of both equations we obtain for the supply and demand equations (using Δ^j to denote a time and reference country difference of a variable)

$$\begin{aligned}
(1+\omega_2)\,\Delta^j \ln uv_{it}^{k} - \omega_2 \Delta^j \ln X_{it}^{k} &= \Delta^j \tilde{\xi}_{it}^{k} \\
\ln \Delta^j X_{it}^{k} + (\sigma-1)\,\Delta^j \ln uv_{it}^{k} &= \Delta^j \tilde{\epsilon}_{it}^{k}
\end{aligned}$$

and these are precisely the supply and demand equations from Feenstra (1994). Then following the same steps as Feenstra (1994) we can obtain the analogous estimating equation.

V Comparing Feenstra (1994) and Feenstra and Romalis (2014) estimates

In this section of the Appendix, compare estimates from Feenstra (1994) (conditional on passing vs failing the tests) to those from Feenstra and Romalis (2014).

First, we explain how we compare parameter estimates from Feenstra (1994) to thos from Feenstra and Romalis (2014). Direct parameter comparisons of σ and ω across the two models are misleading because these parameters capture different things in the two models. In Feenstra and Romalis (2014) demand and supply elasticities are convoluted by firm heterogeneity and quality. However, the parameter governing vertical shifts in supply represents the same fundamental mechanism across both models. We show Feenstra and Romalis (2014) implies the parameter[41] $\rho_{FR} \equiv \frac{\omega_{FR}(\sigma_{FR}-1)}{1+\omega_{FR}\left(1+\frac{(\sigma_{FR}-1)(1+\gamma_{FR})}{1+\theta_{FR}(\sigma_{FR}-1)}\right)}$ governs shifts in their supply curves, and is directly comparable to Feenstra (1994)'s $\rho \equiv \frac{\omega(\sigma-1)}{1+\omega\sigma}$. In order to make comparisons across supply and demand elasticity estimates we define the productivity distribution adjusted supply elasticity $S_{FR} \equiv \frac{\omega_{FR}}{1+\omega_{FR}}(1+\gamma)$ and the quality adjusted demand elasticity $D_{FR} \equiv \frac{\sigma_{FR}-1}{1+\theta_{FR}(\sigma_{FR}-1)}$.

In the text, we compare our estimates to those from Feenstra and Romalis (2014) in Table 7 conditional on passing and failing the tests for the overall sample and disaggregated by Rauch Classification. In Table 9, we disaggregate by SITC 1-digit code and show that there are similar patters as in the overall sample.

Table 9: Relation of Tests Disaggregated by SITC1 to Feenstra and Romalis (2014) Estimates

				Feenstra and Romalis (2014)					
		Pass Tests		Supply Shifter		Supply Elasticity		Demand Elasticity	
SITC1	Products	J-test	F-test	$\hat{\rho}_{FR}$	$\frac{\hat{\rho}-\hat{\rho}_{FR}}{1/2(\hat{\rho}_{FR}+\hat{\rho})}$	$\hat{S}_{FR}$	$\frac{\frac{\hat{\omega}}{1+\hat{\omega}}-\hat{S}_{FR}}{1/2\left(\hat{S}_{FR}+\frac{\hat{\omega}}{1+\hat{\omega}}\right)}$	$\hat{D}_{FR}$	$\frac{\hat{\sigma}-\hat{D}_{FR}}{1/2(\hat{D}_{FR}+\hat{\sigma})}$
0	48	19%	71%	0.165	−0.080	0.379	−0.561	2.331	0.184
0	29	100%	100%	0.160	−0.070	0.457	−0.710	2.341	−0.123
1	6	0%	100%	0.284	−1.512	3.226	−1.819	2.369	0.238
1	5	100%	100%	0.333	−0.130	2.701	−1.492	2.744	0.650
2	46	50%	41%	0.207	−0.709	0.792	−1.205	2.224	0.065
2	27	100%	100%	0.208	−0.008	1.118	−0.929	2.062	0.224
3	5	40%	40%	0.051	1.774	0.003	1.275	2.399	1.564
3	10	100%	100%	0.024	0.653	0.000	0.566	2.322	0.452
4	8	50%	50%	0.035	1.021	0.002	0.402	2.171	0.784
4	2	100%	100%	0.021	1.176	0.095	0.912	1.977	−0.658
5	57	11%	74%	0.192	−0.464	1.250	−0.969	2.084	0.547
5	32	100%	100%	0.198	−0.188	1.347	−0.975	2.104	−0.189
6	105	25%	70%	0.178	−0.565	0.738	−0.991	2.192	0.354
6	70	100%	100%	0.175	−0.124	0.682	−0.513	2.185	−0.059
7	56	32%	54%	0.237	−0.956	0.900	−1.366	2.296	0.224
7	71	100%	100%	0.229	−0.342	1.049	−0.912	2.300	0.248
8	35	29%	57%	0.231	−0.483	1.062	−1.055	2.318	0.281
8	41	100%	100%	0.230	−0.044	0.977	−0.614	2.294	0.074
9	3	33%	67%	0.275	−0.030	1.164	−0.807	2.310	1.168

Notes: Median estimates of $\rho_{FR} \equiv \frac{\omega_{FR}(\sigma_{FR}-1)}{1+\omega_{FR}\left(1+\frac{(\sigma_{FR}-1)(1+\gamma_{FR})}{1+\theta_{FR}(\sigma_{FR}-1)}\right)}$, $S_{FR} \equiv \frac{\omega_{FR}}{1+\omega_{FR}}(1+\gamma)$, and $D_{FR} \equiv \frac{\sigma_{FR}-1}{1+\theta_{FR}*(\sigma_{FR}-1)}$ are presented in Columns 5,7,9. Differences between Feenstra and Romalis (2014) and our estimates are presented in Columns 6,8,10. For the sake of space, robust standard errors are available with the published estimates.

[41] All of the following notation is defined in prior sections of the Appendix.