Universal Gravity

Treb Allen

Dartmouth College and National Bureau of Economic Research

Costas Arkolakis

Yale University and National Bureau of Economic Research

Yuta Takahashi

Hitotsubashi University

We study the theoretical properties and counterfactual predictions of a large class of general equilibrium trade and economic geography models. By combining aggregate factor supply and demand functions with market-clearing conditions, we prove that existence, uniqueness, and—given observed trade flows—the counterfactual predictions of any model within this class depend only on the demand and supply elasticities ("gravity constants"). Using a new "model-implied" instrumental variables approach, we estimate these gravity constants and use these estimates to compute the impact of a trade war between the United States and China.

I. Introduction

Over the past 15 years, there has been a quantitative revolution in spatial economics. The proliferation of general equilibrium gravity models incorporating flexible linkages across many locations now gives researchers

We thank Andy Atkeson, David Atkin, Lorenzo Calends, Arnaud Costinot, Jonathan Dingel, Dave Donaldson, Jonathan Eaton, Pablo Fajgelbaum, John Geanakoplos, Penny Goldberg, Sam Kortum, Xiangliang Li, Giovanni Maggi, Kiminori Matsuyama, Francesc Ortega, Ralph

Electronically published December 23, 2019
[*Journal of Political Economy*, 2020, vol. 128, no. 2]

the ability to conduct a rich set of real-world analyses. However, the complex general equilibrium interactions and the variegated assumptions underpinning different models have had the result that our understanding of the models' properties lags behind. As a result, many important questions remain either partially or fully unresolved, including "When does an equilibrium exists and when is it unique?" and "Do different models have different counterfactual implications?"

In this paper, we characterize the theoretical and empirical properties common to a large class of gravity models spanning the fields of international trade and economic geography. We first provide a "universal gravity" framework, combining aggregate-demand and aggregate-supply equations with standard market-clearing conditions, that incorporates many workhorse trade and economic geography models.[1] We show that existence and uniqueness of the equilibria of all models under the auspices of our framework can be characterized solely on the basis of their aggregate-demand and aggregate-supply elasticities (the "gravity constants"). Moreover, the counterfactual predictions for trade flows, incomes, and prices of these models can be expressed solely as functions of the gravity constants and observed data. Hence, the key theoretical properties and positive counterfactual predictions of all gravity models depend ultimately on the value of two parameters: the elasticities of supply and demand. We show how these gravity constants can be estimated by using a new instrumental variables approach that exploits the general equilibrium structure of the model. Finally, we use these estimates to compute the impact of a trade war between the United States and China.

To construct our framework, we consider a representative economy in which an aggregate good is traded across locations subject to the following six economic conditions: (1) "iceberg"-type bilateral trade frictions; (2) a constant elasticity of substitution (CES) aggregate-demand function; (3) a CES aggregate-supply function; (4) market clearing; (5) exogenous trade deficit; and (6) a choice of the numeraire. Any model in which the equilibrium can be represented in a way that satisfies these conditions

Ossa, Nina Pavcnik, Steve Redding, Andres Rodriguez-Clare, Bob Staiger, Chris Tonetti, our editor Harald Uhlig, and four anonymous referees for excellent comments and suggestions. This material is based upon work supported by the National Science Foundation under grants SES-1658838 and SES-1658875. A Matlab tool kit that is the companion to this paper is available on Allen's website. All errors are our own. Data are provided as supplementary material online.

[1] Examples of gravity trade models included in our framework are perfect-competition models such as those of Anderson (1979), Eaton and Kortum (2002), Anderson and van Wincoop (2003), Dekle, Eaton, and Kortum (2008), and Caliendo and Parro (2015); monopolistic competition models such as those of Krugman (1980), Melitz (2003) as specified by Arkolakis et al. (2008), Chaney (2008), and Di Giovanni and Levchenko (2010); and the Bertrand competition model of Bernard et al. (2003). Economic geography models incorporated in our framework include those of Allen and Arkolakis (2014) and Redding (2016). See table 1 for the mapping from workhorse trade and economic geography models into the universal gravity framework.

is said to be contained within the universal gravity framework. Moreover, these conditions impose sufficient structure to completely characterize all general equilibrium interactions of trade flows, incomes, and prices. The aggregate-demand elasticity from condition 2 and the aggregate-supply elasticity from condition 3 play a particularly important role in this characterization.

We first provide sufficient conditions for the existence, uniqueness, and interiority of the equilibrium of the model that depend solely on the gravity constants. Existence occurs everywhere except for a knife-edge constellation of parameters (corresponding, e.g., to Leontief preferences in an Armington trade model or when agglomeration forces are just strong enough to create a "black-hole" equilibrium in an economic geography model). An equilibrium is unique as long as the demand elasticity is (weakly) negative and the supply elasticity is (weakly) positive (or vice versa and both elasticities are greater than one in magnitude). Multiplicity may occur if demand and supply elasticities are both negative (e.g., in an economic geography model if agglomeration forces are sufficiently strong) or if demand and supply elasticities are both positive (e.g., in a trade model if goods are complementary). We also show that these sufficient conditions can be extended further if trade frictions are "quasi-symmetric"—a common assumption in the literature.

We then examine how a shock to bilateral trade frictions affects equilibrium trade flows, incomes, and prices. To do so, we derive an analytical expression for the counterfactual elasticities of these endogenous variables to changes in all bilateral trade frictions that elucidates the network effects of trade. In particular, we show how this expression can be written as series of terms expressing how a shock propagates through the trading network, for example, the direct effect of a shock, the effect of the shock on all locations' trading partners, the effect on all locations' trading partners' trading partners, and so on. Importantly, we show that this expression depends only on observed trade flows and the gravity constants, demonstrating that, conditional on these two model parameters, the positive macroeconomic implications for all gravity models are the same.[2] Moreover, we analytically prove that when trade frictions are "quasi-symmetric," the impact of a trade friction shock on the real output prices and real expenditure in directly affected locations will always exceed the impact on other, indirectly affected locations.

We proceed by estimating the gravity constants, using a procedure that can be applied to any model contained within the universal gravity framework. We show that the supply and demand elasticities can be estimated by regressing a location's fixed effect (recovered from a gravity equation)

[2] While the implications for real output prices are the same for all gravity models, the mapping from real output prices to welfare will, in general, depend on the particular model. As a result, the normative (welfare) implications will vary across different models, as we discuss in detail below.

on its own expenditure share (the coefficient of which is the supply elasticity) and its income (the coefficient of which is the demand elasticity). Identifying the elasticities requires a set of instruments that are correlated with own expenditure share and income but uncorrelated with unobserved supply shifters (such as productivity) in the residual. We construct "model-implied" instruments that use the general equilibrium structure of the model by calculating the equilibrium own expenditure shares and incomes of a hypothetical world where no such unobserved supply shifters exist and bilateral trade frictions are only a function of distance. Using this procedure, we estimate a demand elasticity in line with previous estimates from the trade literature (e.g., Simonovska and Waugh 2014) and a supply elasticity that is larger than is typically calibrated to in trade models but appears reasonable given estimates from the economic geography literature.

Finally, we use the estimated gravity constants, along with the expression for comparative statics, to evaluate the effect of a trade war between the United States and China on the real expenditure of all countries in the world. Given our large estimated supply elasticity, we find modest declines in (real) prices but large declines in (real) expenditure. Third-country effects are also substantial, with important trading partners of China (e.g., Vietnam and Japan) and the United States (e.g., Canada and Mexico) being especially adversely affected.

This paper is related to a number of strands of literature in the fields of international trade, economic geography, and general equilibrium theory. There is a small but growing literature examining the structure of general equilibrium models of trade and economic geography. In particular, Arkolakis, Costinot, and Rodríguez-Clare (2012) provide conditions under which a model yields a closed-form expression for changes in welfare as a function of changes in openness, while in a recent paper Adao, Costinot, and Donaldson (2017) show how to conduct counterfactual predictions in neoclassical trade models without imposing gravity. In contrast, our paper incorporates models with elastic aggregate-supply curves, thereby allowing analysis of both economic geography models and trade models with intermediate "roundabout" production. A key characteristic of the class of models we study is that the "gravity constants" are the same across all locations; while strong, this assumption imposes sufficient structure to completely characterize all general equilibrium interactions while retaining tractability, even in the presence of a large number of locations.[3]

In terms of the theoretical properties of the equilibrium, Alvarez and Lucas (2007) use the gross-substitutes property to establish sufficient

[3] In contrast, the literature on computable general equilibrium models typically focuses on models with a large number of elasticities (e.g., location or region specific) but only a small number of regions; for a review of these models, see Menezes et al. (2006). Although outside the purview of this paper, it would be perhaps be interesting future work to determine whether some of the tools developed below could be applied to those models.

conditions for uniqueness for gravity trade models. We instead generalize results from the study of nonlinear integral equations (see, e.g., Karlin and Nirenberg 1967; Zabreyko et al. 1975; Polyanin and Manzhirov 2008) to systems of nonlinear integral equations. As a result, the sufficient conditions we provide are strictly weaker than those derived by Alvarez and Lucas (2007). In particular, our conditions allow the supply elasticity to be larger in magnitude than the demand elasticity (in which case gross substitutes may not hold), which is what we find when we estimate the elasticities. In previous work, Allen and Arkolakis (2014) provide sufficient conditions for existence and uniqueness for economic geography models. Unlike those results, our conditions do not require symmetric trade frictions, nor do we require finite trade frictions between all locations. Unlike both Alvarez and Lucas (2007) and Allen and Arkolakis (2014), our theoretical results cover both trade and economic geography models simultaneously.

Our analytical characterization of the counterfactual predictions is related to the "exact hat algebra" methodology pioneered by Dekle, Eaton, and Kortum (2008) and extended by Costinot and Rodríguez-Clare (2013) and many others. Unlike that approach, we characterize the elasticity of endogenous variables to trade shocks (i.e., we examine local shocks instead of global shocks). There are several advantages of our local approach. First, all possible counterfactuals can be calculated simultaneously through a single matrix inversion. Second, our analytical characterization holds for local shocks around the observed equilibria even if there are other possible equilibria (in which case we are unaware of a procedure that ensures the solution to the "exact hat" approach that corresponds to the observed equilibria). Third, the local analytical expression admits a simple economic interpretation as a shock propagating through the trading network. In this regard, our paper is related to the recent working paper by Bosker and Westbrock (2016), which examines how shocks propagate through global production networks. Fourth, our analytical derivation allows us to characterize the relative size of the elasticity of real output prices and real output in different locations from a trade friction shock, providing (to our knowledge) one of the first analytical results about the relative size of the direct and indirect impacts of a trade friction shock in a model with many locations.

Our estimation strategy uses equilibrium income and own expenditure shares from a hypothetical economy as instruments to identify the demand and supply elasticities. Following Eaton and Kortum (2002), we use the fixed effects of a gravity equation as the dependent variable in an instrumental variables regression (although we use the regression to estimate the supply elasticity along with the demand elasticity). One advantage of our approach is the simplicity of calculating our instruments by using bilateral distances and observed geographic variables; in this regard, we owe credit to Frankel and Romer (1999), who instrument for trade with geography (albeit not in a general equilibrium context).

The idea of using the general equilibrium structure of the gravity model to recover key parameters is originally due to Anderson and van Wincoop (2003). Following this, several papers have sought to improve the typical gravity equation estimation by accounting for equilibrium conditions. For example, Anderson and Yotov (2010) pursue an estimation strategy, imposing that the equilibrium "adding-up constraints" of the multilateral resistance terms are satisfied, whereas Fally (2015) proposes the use of a Poisson pseudo–maximum likelihood estimator whose fixed effects ensure that such constraints are satisfied, and Egger and Nigai (2015) develop a two-step model consistent approach that overcomes bias arising from general equilibrium forces and unobserved trade frictions. Unlike these papers, here our focus is on recovering the demand and supply elasticities rather than estimating trade friction coefficients in a model consistent manner.

Recent work by Anderson, Larch, and Yotov (2015) explores the relationship between trade and growth examined by Frankel and Romer (1999) in a structural context. They recover the demand (trade) elasticity from a regression of income on a multilateral resistance term, where endogeneity concerns are addressed by calculating multilateral resistance based on international linkages only. Our estimation strategy, in contrast, recovers both the demand and supply elasticities from a gravity regression and overcomes endogeneity concerns, using a new instrumental variables approach based on the general equilibrium structure of the model.

Finally, we should note that the brief literature review above is by no means complete and refer the interested reader to the excellent review articles by Baldwin and Taglioni (2006), Head and Mayer (2013), Costinot and Rodríguez-Clare (2013), and Redding and Rossi-Hansberg (2017); the latter two focus especially on quantitative spatial models.

The remainder of the paper is organized as follows. In the next section, we present the universal framework and discuss how it nests existing general equilibrium gravity models. In section III, we present the theoretical results for existence and uniqueness. In section IV, we present the results concerning the counterfactual predictions of the model. In section V, we estimate the gravity constants. In section VI, we calculate the effects of a United States–China trade war. Section VII concludes.

II. A Universal Gravity Framework

Before turning to the universal gravity framework, we present two variants of the simple Armington gravity model to provide a concrete example of the type of models that fall within our framework. Suppose that there are N locations, each producing a differentiated good, and in what follows we define the set $S \equiv \{1, \ldots, N\}$. The only factor of production is labor, where we denote the allocation of labor in location $i \in S$ as L_i and assume that the total world labor endowment is $\Sigma_{i \in S} L_i = \bar{L}$. Shipping the good from $i \in S$ to $j \in S$ incurs an *iceberg trade friction*, where $\tau_{ij} \geq 1$ units must be shipped in order for one unit to arrive. Consumers have CES preferences with elasticity of substitution $\sigma \geq 0$.

In the first variant, which we call the "trade" model, suppose that the labor endowed to a location is exogenous and perfectly inelastic, as in Anderson (1979) and Anderson and van Wincoop (2003). Suppose too that there is roundabout production, as in Eaton and Kortum (2002), that combines labor and an intermediate input in a Cobb-Douglas fashion. Thus, the quantity of output produced in location i is $Q_i = (A_i L_i)^{\zeta} I_i^{1-\zeta}$, with $\zeta \in (0, 1]$ the labor share, A_i is the labor productivity in location $i \in S$, and I_i is an intermediate input equal to a CES aggregate of the differentiated varieties in all locations with the same elasticity of substitution σ as final demand. In this case, the output price in location i is $p_i = (w_i/A_i)^{\zeta} P_i^{1-\zeta}$, where w_i is the wage and $P_j \equiv [\Sigma_{k \in S}(p_k \tau_{kj})^{1-\sigma}]^{1/(1-\sigma)}$ is both the CES price index for the consumer and the price per unit of intermediate input.

In the second variant, the "economic geography" model, we suppose instead that the labor supplied to a location is perfectly elastic, so that welfare is equalized across locations, as in Allen and Arkolakis (2014).[4] Welfare in this model is the product of the real expenditure of labor and the amenity value of living in a location, denoted by u_i, and welfare equalization implies $(w_i/P_i)u_i = (w_j/P_j)u_j$ for all $i, j \in S$. We further assume that productivities and amenities are subject to spillovers: $A_i = \bar{A}_i L_i^a$ and $u_i = \bar{u}_i L_i^b$. In this variant of the model, the quantity of output produced in location i is $Q_i = \bar{A}_i L_i^{1+a}$, and the output price is $p_i = w_i/(\bar{A}_i L_i^a)$.[5]

In both variants of the model, CES consumer preferences for the goods from each location yield a gravity equation that characterizes the aggregate demand in location j for the differentiated variety from location i:

$$X_{ij} = \frac{(p_i \tau_{ij})^{1-\sigma}}{\sum_{k \in S} (p_k \tau_{kj})^{1-\sigma}} E_j, \tag{1}$$

for all j, where $E_j = \Sigma_{j \in S} X_{ji}$ is the expenditure in location j.

[4] In addition, this formulation incorporates many prominent economic geography models, e.g., those of Helpman (1998), Donaldson and Hornbeck (2012), Bartelme (2014), and Redding (2016).

[5] It is straightforward to add roundabout production into the economic geography variant of the model (see table 1); we omit doing so here to keep our illustrative examples as simple as possible.

More subtly, both variants of the model also feature an aggregate supply for the quantity of output produced in each location. In the trade variant of the model—despite the labor supply being perfectly inelastic—we can use the fact that a constant share of revenue is paid to both workers and intermediates to write the output of location i as

$$Q_i = A_i L_i \left(\frac{p_i}{P_i}\right)^{(1-\zeta)/\zeta}. \tag{2}$$

Similarly, in the economic geography variant of the model we can use the welfare equalization condition to write:

$$Q_i = \kappa \bar{A}_i^{(b-1)/(a+b)} \bar{u}_i^{-[(1+a)/(a+b)]} \left(\frac{p_i}{P_i}\right)^{-[(1+a)/(a+b)]}, \tag{3}$$

where $\kappa \equiv \{\bar{L}/[\Sigma_{i\in S}(\bar{A}_i\bar{u}_i)^{-[1/(a+b)]}(p_i/P_i)^{-[1/(a+b)]}]\}^{1+a}$ is an (endogenous) scalar that depends on the aggregate labor endowment $\bar{L}$ and we refer to p_i/P_i as the *real output price* in location $i \in S$.[6] Finally, in both variants, we close the model by requiring that the value of total output equals total sales (market clearing), that is,

$$Y_i \equiv p_i Q_i = \sum_{j\in S} X_{ij}, \tag{4}$$

and that total expenditure equals total output (balanced trade), that is,

$$E_i = p_i Q_i. \tag{5}$$

Substituting the CES demand (eq. [1]) and supply (eqq. [2] or [3]) equations into the market-clearing and balanced-trade conditions yields the following identical system of equilibrium equations for both variants of the model. In particular,

$$p_i^{1+\phi}\bar{c}_i\left(\frac{p_i}{P_i}\right)^{\psi} = \sum_{j\in S}\tau_{ij}^{-\phi}P_j^{\phi}p_j\bar{c}_j\left(\frac{p_j}{P_j}\right)^{\psi} \ \forall\ i \in S, \tag{6}$$

$$P_i^{-\phi} = \sum_{j\in S}\tau_{ji}^{-\phi}p_j^{-\phi} \ \forall\ i \in S, \tag{7}$$

where in the trade variant of the model $\psi \equiv (1-\zeta)/\zeta$ and $\bar{c}_i \equiv A_i L_i$, in the economic geography variant of the model $\psi \equiv -[(1+a)/(a+b)]$ and $\bar{c}_i \equiv \bar{A}_i^{(b-1)/(a+b)}\bar{u}_i^{-[(1+a)/(a+b)]}$, and in both models $\phi \equiv \sigma - 1$. Note that

[6] In these two examples—as in most of the analysis that follows—we focus on interior equilibria where production is positive in all locations. In app. B.2 (app. B is available online), we generalize our setup to allow for the possibility of noninterior solutions where production is zero in some locations, which allows, e.g., for the case that welfare in unpopulated locations may be lower than that in populated locations. In theorem 1 below, we provide sufficient conditions under which all equilibria are guaranteed to be interior.

in both models the constants $\{\bar{c}_i\}_{i\in S}$ are exogenous model location-specific fundamentals, which we refer to as *supply shifters* in what follows, and ϕ and ψ are parameters. Given supply shifters, trade frictions, and the two parameters, one can use equations (6) and (7) to solve for output prices p_i and price indices P_i (up to scale). One can then use a normalization that total world income is equal to one, that is, $\Sigma_{i\in S} Y_i = 1$ and the gravity equation (eq. [1]) to calculate trade flows X_{ij}. Given trade flows, income Y_i can then be recovered from market clearing (eq. [4]). Note that although the endogenous scalar κ from the economic geography model does not enter the equilibrium system of equations (and hence does not affect trade flows or incomes), it does affect the level of output, a point we return to below.

This example highlights the close relationship between trade and geography models and suggests the possibility for a unified analysis of the properties of such spatial gravity models. In what follows, we present a framework comprising six simple economic conditions about aggregate trade flows of a representative good between many locations. We show that the equilibrium of any model that satisfies these conditions can be represented by the solution to equations (6) and (7).

To proceed with our universal gravity framework, it is helpful to first introduce some terminology. Define the *output* $Q_i \geq 0$ to be the quantity of the representative good produced in location $i \in S$; the *quantity traded* $Q_{ij} \geq 0$ be the quantity of the representative good in location $i \in S$ that is consumed in location $j \in S$; the *output price* $p_i \geq 0$ to be the (factory-gate) price per unit of the representative good in location $i \in S$; the *bilateral price* $p_{ij} \geq 0$ to be the cost of the representative good from location $i \in S$ in location $j \in S$; the *income* $Y_i \equiv p_i Q_i$ to be the total value of the representative good in location $i \in S$; the *trade flows* $X_{ij} \equiv p_{ij} Q_{ij}$ to be the value of the good in $i \in S$ sold to $j \in S$; the *expenditure* $E_i \equiv \Sigma_{j\in S} X_{ji}$ to be the total value of imports in $i \in S$; the *real expenditure* $W_i \equiv E_i/P_i$ to be a measure of expenditure in location $i \in S$, where P_i is a *price index* defined below; and the *real output price* (referred to simply as "prices" in sec. I) to be p_i/P_i.[7]

We say that an equilibrium is *interior* if output and output prices are strictly positive in all locations, that is, $Q_i > 0$ and $p_i > 0$ for all $i \in S$. In what follows, we focus our attention on interior equilibria and disregard the trivial equilibrium where $Q_i = 0$ for all $i \in S$. We provide sufficient conditions to ensure that all equilibria are interior below and examine noninterior solutions in depth in appendix B.2. Clearly, because of the presence of complementarities there is a possibility of multiple

[7] Because the real output price is the ratio of the price of goods sold to the price index of goods purchased, it is closely related to the terms of trade, which is the ratio of export prices to import prices, differing only in that the price index also includes goods purchased domestically.

interior equilibria. This is true in the economic geography model, because of labor mobility and agglomeration externalities, or even in the trade model, when complementarities in consumption are large (low σ).

We first start with a condition that describes the relationship between the output price in location i and the bilateral price:

CONDITION 1. The bilateral price is equal to the product of the output price and a bilateral scalar:

$$p_{ij} = p_i \tau_{ij}, \tag{8}$$

where, as above, $\{\tau_{ij}\}_{i,j \in S} \in \bar{\mathbb{R}}_{++}$ are referred to as "trade frictions."[8]

Given prices, the next condition can be used to derive aggregate demand.

CONDITION 2 (CES aggregate demand). There exists an exogenous (negative of the) demand elasticity $\phi \in \mathbb{R}$ such that the expenditure in location $j \in S$ can be written as

$$E_j = \left(\sum_{i \in S} p_{ij}^{-\phi} \right)^{-(1/\phi)} W_j, \tag{9}$$

where W_j is the real expenditure and the associated price index is $P_j \equiv (\sum_{i \in S} p_{ij}^{-\phi})^{-(1/\phi)}$. By Shephard's lemma, condition 2 (hereafter C.2) implies that the trade flows from $i \in S$ to $j \in S$ can be written as

$$X_{ij} = \frac{p_{ij}^{-\phi}}{\sum_{k \in S} p_{kj}^{-\phi}} E_j. \tag{10}$$

We refer to equation (10) as the *aggregate demand* of the universal gravity model. The aggregate-demand equation (10), combined with C.1, yields a *gravity equation* equivalent to equation (2) in Anderson and van Wincoop (2004), condition R3′ in Arkolakis, Costinot, and Rodríguez-Clare (2012), and the CES factor demand specification considered in Adao, Costinot, and Donaldson (2017). Accordingly, we note that the demand elasticity ϕ is often referred to as the "trade elasticity" in the literature.

It is important to emphasize that real expenditure $W_i = E_i/P_i$ and real output prices p_i/P_i are concepts distinct from welfare, as neither necessarily corresponds to the welfare of the underlying factor of production (such as labor) of a particular model. In the models above, for example, the welfare of a worker corresponds to her real wage, which is equal to the marginal product of a worker divided by the price index. Because of the presence of roundabout production (in the trade model) or externalities (in the economic geography model), a worker's marginal product is not equal to the price per unit (gross) output.[9]

[8] We define $\bar{\mathbb{R}}_{++}$ as $\mathbb{R}_{++} \cup \{\infty\}$. If $\tau_{ij} = \infty$, then there is no trade between i and j.

[9] The relationship between real output prices and welfare for a number of seminal models is summarized in the last column of table 1 and discussed in detail in app. B.11.

We furthermore assume that output in a location is potentially endogenous and specify the following supply-side equation.

CONDITION 3 (CES aggregate supply). There exist exogenous supply shifters $\{\bar{c}_i\} \in \mathbb{R}^N_{++}$, an exogenous aggregate-supply elasticity $\psi \in \mathbb{R}$, and an endogenous scalar $\kappa > 0$ such that output in each location $i \in S$ can be written as

$$Q_i = \kappa \bar{c}_i \left(\frac{p_i}{P_i}\right)^{\psi}. \tag{11}$$

In what follows, we refer to equation (11) as the *aggregate supply* of the universal gravity model and the set of demand and supply elasticities (ϕ, ψ) as the *gravity constants*.

In general, the value of the endogenous scalar κ will depend on the particular model; for example, as we saw above, in the trade model $\kappa = 1$, whereas in the economic geography model κ is endogenously determined. Without taking a particular stance on the underlying model (and the implied value of κ), the scale of output is unspecified.[10] However, we show below that we can still identify the equilibrium trade flows, incomes, and real output prices—including their level—without knowledge of κ.

Finally, to close the model, we impose two standard conditions and choose our numeraire.

CONDITION 4 (Output market clearing). For all $i \in S$, $Q_i = \sum_{j\in S}\tau_{ij}Q_{ij}$. Note that by multiplying both sides of C.4 by the output price, we have that income is equal to total sales, as in equation (4) in our example economy.[11]

CONDITION 5 (Exogenous deficits). For all $i \in S$, $E_i = \Xi\xi_i p_i Q_i$, where ξ_i is the exogenous expenditure-output ratio for location i up to constant and Ξ is an endogenous scalar that ensures that the world market-clearing condition holds:

$$\Xi = \frac{\sum_i p_i Q_i}{\sum_i \xi_i p_i Q_i}. \tag{12}$$

We say that trade is *balanced* in the special case that $\xi_i = 1$ for all $i \in S$ (in which case $\Xi = 1$). While balanced trade is a standard assumption in (static) gravity models, we allow for (exogenous) trade imbalances in order to match observed trade data.

Our final condition is a normalization.

[10] While one can choose units of output to ensure that $\kappa = 1$ in any given equilibrium, changes in model fundamentals, given this choice of units, will generally result in κ varying.

[11] As Anderson and van Wincoop (2004) show, one can combine C.1, C.2, and C.4 to derive a gravity equation of the form $X_{ij} = (\tau_{ij}/\Pi_i P_j)^{-\phi} Y_i E_j$, where $\prod_i^{-\phi} \equiv \sum_{j\in S}(\tau_{ij}/P_j)^{-\phi}E_j$ and $P_j^{-\phi} \equiv \sum_{i\in S}(\tau_{ij}/\Pi_i)^{-\phi}Y_i$ are outward and inward multilateral resistance terms, respectively.

Condition 6. World income equals one:

$$\sum_i Y_i = 1. \tag{13}$$

In the absence of a normalization, the level of prices are undetermined because equations (6) and (7) are homogeneous of degree 0 in $\{p_i, P_i\}_{i\in S}$. Moreover, without specifying κ in equation (11), the level of output is also unknown. The choice of normalizing world income to one in C.6 addresses both these issues simultaneously by pinning down the product of the level of these two unknown scalars. As a result, we can determine the equilibrium level (i.e., including scale) of nominal incomes and trade flows. However, the cost of doing so is that both the level of output (in quantities) and prices remain unknown. As a result, the primary focus in the following analysis is on three endogenous model outcomes for which we can pin down the levels: incomes, trade flows, and real output prices $\{p_i/P_i\}_{i\in S}$ (which are invariant to both κ and the scale of prices and hence determined including scale).

Given any gravity constants $\{\phi, \psi\}$, supply shifters, $\{\bar{c}_i\}_{i\in S}$, and bilateral trade frictions $\{\tau_{ij}\}_{i,j\in S}$, we define an *equilibrium of the universal gravity framework* to be a set of endogenous outcomes determined up to scale—namely, outputs $\{Q_i\}_{i\in S}$, quantities traded $\{Q_{ij}\}_{i,j\in S}$, output prices $\{p_i\}_{i\in S}$, bilateral prices $\{p_{ij}\}_{i,j\in S}$, price indices $\{P_i\}_{i\in S}$, and real expenditures, as well as a set of endogenous outcomes for which the scale is known, namely, incomes $\{Y_i\}_{i\in S}$, expenditures $\{E_i\}_{i\in S}$, trade flows $\{X_{ij}\}_{i,j\in S}$, and real output prices $\{p_i/P_i\}_{i\in S}$—that together satisfy C.2–C.6.

As table 1 summarizes, many well-known trade and economic geography models are contained within the universal gravity framework. On the demand side, it is well known (see, e.g., Adao, Costinot, and Donaldson 2017 and Arkolakis et al. 2019) that many trade models imply an aggregate-CES-demand system, as specified in C.2.[12] For example, in the Armington perfect-competition model, a CES demand, combined with linear production functions, implies that $\phi = \sigma - 1$; in the Eaton and Kortum (2002) model, a Ricardian model with endogenous comparative advantage across goods and Fréchet-distributed productivities across sectors with elasticity θ implies that $\phi = \theta$. Similarly, a class of monopolistic models with CES or non-CES demand, linear production function, and Pareto-distributed productivities with elasticity θ, summarized in Arkolakis et al. (2019), also implies $\phi = \theta$. Economic geography models delivering gravity equations for trade flows, such as those of Allen and Arkolakis (2014) and Redding (2016), also satisfy C.2.

[12] The class of trade models considered by Arkolakis, Costinot, and Rodríguez-Clare (2012), under their CES demand assumption R3′, are a strict subset of the models that fall within the universal gravity framework, corresponding to the case of $\psi = 0$.

As discussed in the example above, labor mobility across locations generates a CES aggregate supply satisfying C.3, with a supply elasticity of $\psi = -[(1+a)/(a+b)]$. In this case, the supply elasticity depends on the strength of the agglomeration/dispersion forces summarized by $a+b$. Assuming that $a > -1$, if dispersion forces dominate $(a+b<0)$, the supply elasticity is positive, whereas when agglomeration forces dominate $(a+b<0)$, the supply elasticity is negative.

Perhaps more surprising, trade models incorporating "roundabout" trade with intermediate goods also exhibit an aggregate CES supply, even though workers are immobile across locations. As discussed in the example above, the supply elasticity is $\psi = (1-\zeta)/\zeta$ and hence positive and increasing in the share of intermediates in the production. In the next two sections, we show that any trade and economic geography models sharing the same gravity constants will also share the same theoretical properties and counterfactual implications.

What types of models are not contained within the universal gravity framework? Conditions 2 and 3 are violated by models that do not exhibit constant demand and supply elasticities, which include those of Novy (2010), Fajgelbaum and Khandelwal (2014), Head, Mayer, and Thoenig (2014), Melitz and Redding (2015), and Adao, Costinot, and Donaldson (2017). Models with multiple factors of production with nonconstant factor intensities will generally not admit a single aggregate-good representation and hence are also not contained within the universal gravity framework (although the tools developed below can often be extended to analyze such models, depending on the particular functional forms). Condition 5 is violated both by dynamic models in which the trade deficits are endogenously determined and by models incorporating additional sources of revenue (such as tariffs); hence, these models are not contained within the universal gravity framework. However, we show in appendix B.8 how the results below can be applied to a simple Armington trade model with tariffs.[13] Finally, while the universal gravity framework includes a single sector, the mathematical tools used to prove existence and uniqueness below can be extended to allow for multiple sectors of production as in, for example, Costinot, Donaldson, and Komunjer (2012); see Allen, Arkolakis, and Li (2014).

[13] It is important to note that while the universal gravity framework can admit tariffs, how tariffs affect the model implications will in general depend on the microeconomic foundations of a model. In particular, the Armington model presented in app. B.8 abstracts from two additional complications that may arise with the introduction of tariffs. First, the elasticity of trade to tariffs may be different from the elasticity of trade to trade frictions, depending on the model; second, if one does not impose that tariffs are uniform for all trade flows between country pairs, the construction of (good-varying) optimal tariffs will depend on the particular microeconomic structure of the model; see Costinot, Rodríguez-Clare, and Werning (2016) for a detailed discussion of these issues.

TABLE 1
EXAMPLES OF MODELS IN THE UNIVERSAL GRAVITY FRAMEWORK

Model	Demand Elasticity (ϕ)	Supply Elasticity (ψ)	Model Parameters	Welfare of Labor
Armington 1969; Anderson 1979; Anderson and van Wincoop 2003 (with intermediates)	$\sigma - 1$	$(1-\zeta)/\zeta$	σ, substitution parameter; ζ, labor share	$B_i \times (p_i/P_i)^{1+\psi}$
Krugman 1980 (with intermediates)	$\sigma - 1$	$(1-\zeta)/\zeta$	σ, substitution parameter; ζ, labor share	$B_i \times (p_i/P_i)^{1+\psi}$
Eaton and Kortum 2002 (with intermediates)	θ	$(1-\zeta)/\zeta$	θ, heterogeneity parameter; ζ, labor share	$B_i \times (p_i/P_i)^{1+\psi}$
Melitz 2003; Di Giovanni and Levchenko 2013	$\theta[1 + \frac{\theta-(\sigma-1)}{\theta(\sigma-1)}]$	$(1-\zeta)/\zeta$	σ, substitution parameter; θ, heterogeneity parameter	$B_i \times (p_i/P_i)^{1+\psi}$
Allen and Arkolakis 2014	$\sigma - 1$	$-[(1+a)/(a+b)]$	σ, substitution parameter; a, productivity spillover; b, amenity spillover	$[\sum_i B_i (p_i/P_i)^{-1/(a+b)}]^{-(a+b)}$
Redding 2016	θ	$\alpha\varepsilon/[1+\varepsilon(1-\alpha)]$	θ, heterogeneity (goods) parameter; ε, heterogeneity (labor) parameter; α, goods expenditure share	$[\sum_i B_i (p_i/P_i)^{\alpha\varepsilon}]^{1/\varepsilon}$

Redding and Sturm 2008 (a variant of Helpman 1998)	$\sigma - 1$	$\alpha/[(1-\alpha)(\sigma-1)-\alpha]$	σ, substitution parameter; α, share spent on goods	$(\sum_i B_i (p_i/P_i)^{\psi\phi})^{(1+\psi)/(1+\psi\phi+\psi)}$
Economic geography with intermediate goods and spillovers	θ	$\frac{1-\zeta}{\zeta} - \frac{1}{\zeta b + a}\left(\frac{a+\zeta}{\zeta}\right)$	θ, heterogeneity (goods) parameter; ζ, intermediate good share; a, productivity spillover; b, amenity spillover	$\sum_i B_i (p_i/P_i)^{1/\zeta}$
Economic geography with intermediate goods, idiosyncratic preferences, and spillovers	θ	$\frac{1-\zeta}{\zeta} + \frac{\varepsilon}{\zeta - \varepsilon(a+\zeta\beta)}\left(\frac{a+\zeta}{\zeta}\right)$	θ, heterogeneity (goods) parameter; ε, heterogeneity (labor) parameter; ζ, intermediate good share; a, productivity spillover; b, amenity spillover	$\{\sum_i B_i [(\frac{p_i}{P_i})^{1/\zeta}]^{\varepsilon/\{1-\varepsilon[(a/\zeta)+b]\}}\}^{\{1-\varepsilon[(a/\zeta)+b]\}/\varepsilon}$

Note.—This table includes a (nonexhaustive) list of trade and economic geography models that can be examined within the universal gravity framework, the mapping of their structural parameters to the gravity constants, and the relationship between the welfare of workers and the real output prices. B_i is an exogenous, location-specific parameter whose interpretation depends on the particular model. λ is an endogenous variable that affects every country simultaneously.

III. Existence, Uniqueness, and Interiority of Equilibria

We proceed by deriving a number of theoretical properties of the equilibria of all models contained within the universal gravity framework.

To begin, we note that we can combine C.1–C.5 to write the equilibrium output prices and price indices (to scale) as the solution to equations (6) and (7). These equations are sufficient to recover the equilibrium level of real output prices and—given the normalization in C.6—the equilibrium level of incomes, expenditures, and trade flows as well as all other endogenous variables up to scale.[14] As a result, equations (6) and (7) (together with the normalization in C.6) are sufficient to characterize the equilibrium of the universal gravity framework.

Before proceeding, we impose two mild conditions on bilateral trade frictions $\{\tau_{ij}\}_{i,j\in S}$:

Assumption 1. The following parameter restrictions hold: (i) $\tau_{ii} < \infty$ for all $i \in S$; and (ii) the graph of the matrix of trade frictions $\{\tau_{ij}\}_{i,j\in S}$ is strongly connected.

The first part of the assumption imposes strictly positive diagonal elements of the matrix of bilateral trade frictions. The second part of the assumption—strong connectivity—requires that there is a sequential path of finite bilateral trade frictions that can link any two locations i and j for any $i \neq j$. This condition has been applied previously in general equilibrium analysis as a condition for existence by McKenzie (1959, 1961) and Arrow and Hahn (1971); as a condition for invertibility by Cheng (1985) and Berry, Gandhi, and Haile (2013); and as a condition for uniqueness by Arrow and Hahn (1971) and Allen (2012). In our case, these two assumptions are the weakest assumptions on the matrix of trade frictions that we can accommodate in order to analyze existence and uniqueness of interior equilibrium.

We mention briefly (but do not need to assume) a third condition. We say that trade frictions are *quasi-symmetric* if there exist a pair of strictly positive vectors $(\tau_i^A, \tau_i^B) \in \mathbb{R}_{++}^{2N}$ such that for any $i, j \in S$, we can write $\tau_{ij} = \tilde{\tau}_{ij}\tau_i^A\tau_j^B$, where $\tilde{\tau}_{ij} = \tilde{\tau}_{ji}$. Quasi symmetry is a common assumption in the literature (see, e.g., Anderson and van Wincoop 2003, Eaton and Kortum 2002, Waugh 2010, and Allen and Arkolakis 2014), and we prove in appendix B.3 that C.1, C.2, C.4, and C.5, taken together, imply that the origin and destination-specific terms in the bilateral trade flow expression are equal up to scale, that is, that $p_i^{-\phi} \propto p_i^{1+\psi}P_i^{\phi-\psi}\bar{c}_i$, which in turn implies that equilibrium trade flows will be symmetric, that is, that $X_{ij} = X_{ji}$ for all $i, j \in S$. The only way that trade can be balanced when trade frictions are quasi-symmetric is to make trade flows bilaterally balanced. As

[14] See app. B.1 for these derivations.

a result, equations (6) and (7) simplify to a single set of equilibrium equations, which allows us to relax the conditions on the following theorem regarding existence and uniqueness.

THEOREM 1. Consider any model contained within the universal gravity framework where trade is balanced (i.e., $\xi_i = 1$ for all $i \in S$) and assumption 1 is satisfied. Then,

i. if $1 + \psi + \phi \neq 0$, then there exists an interior equilibrium;
ii. if $\phi \geq -1$ and $\psi \geq 0$, then all equilibria are interior;
iii. if $\{\phi \geq 0, \psi \geq 0\}$ or $\{\phi \leq -1, \psi \leq -1\}$ (or, if trade frictions are quasi-symmetric and either $\{\phi \geq -1/2, \psi \geq -1/2\}$ or $\{\phi \leq -1/2, \psi \leq -1/2\}$), then there is a unique interior equilibrium;

Proof. See app. A.1 for parts i and iii and app. B.2 for part ii.

A key advantage of theorem 1 is that despite the large dimensionality of the parameter space (N supply shifters $\{\bar{c}_i\}_{i\in S}$ and N^2 trade frictions $\{\tau_{ij}\}_{i,j\in S}$), the conditions are stated only in terms of the gravity constants. Of course, since we provide sufficient conditions, there may be certain parameter constellations, such as particular geographies of trade frictions, where uniqueness may still occur even if the conditions of theorem 1 are not satisfied.[15]

The sufficient conditions for existence, interiority, and uniqueness from theorem 1 are illustrated in figure 1. In the case of existence, standard existence theorems (see, e.g., Mas-Colell, Whinston, and Green 1995) guarantee existence for endowment economies when preferences are strictly convex. This is also true in the universal gravity framework: existence of an interior equilibrium may fail only when $1 + \psi + \phi = 0$, which corresponds to the Armington trade model (without intermediate goods) where $\sigma = 0$, that is, with Leontief preferences that are not strictly convex. Moreover, in the economic geography example above, an interior equilibrium

[15] Theorem 1 extends the existing results on uniqueness of spatial models to a wider parameter range and a broader class of models. In particular, Alvarez and Lucas (2007) provide an alternative approach based on the gross-substitutes property to provide conditions for uniqueness of the Eaton and Kortum (2002) model. In app. B.6, we show that the gross-substitutes property directly applied to our system may fail if the supply elasticity ψ is larger in magnitude than the demand elasticity ϕ, i.e., in ranges $\psi > \phi \geq 0$ or $\psi < \phi \leq -1$. Theorem 1 provides strictly weaker sufficient conditions in that regard. Such parameter constellations are consistent with economic geography models with weak dispersion forces or trade models with large intermediate goods shares. Importantly, in sec. V, we estimate that $\psi > \phi > 0$ empirically.

Theorem 1 generalizes theorem 2 of Allen and Arkolakis (2014) in three ways: (1) it allows for asymmetric trade frictions; (2) it allows for infinite trade frictions between certain locations; and (3) it applies to a larger class of general equilibrium spatial models, including, notably, trade models with inelastic labor supplies (i.e., models in which $\psi = 0$). Theorem 1 also provides a theoretical innovation, as it shows how to extend the mathematical argument of Karlin and Nirenberg (1967) to multiequation systems of nonlinear integral equations.

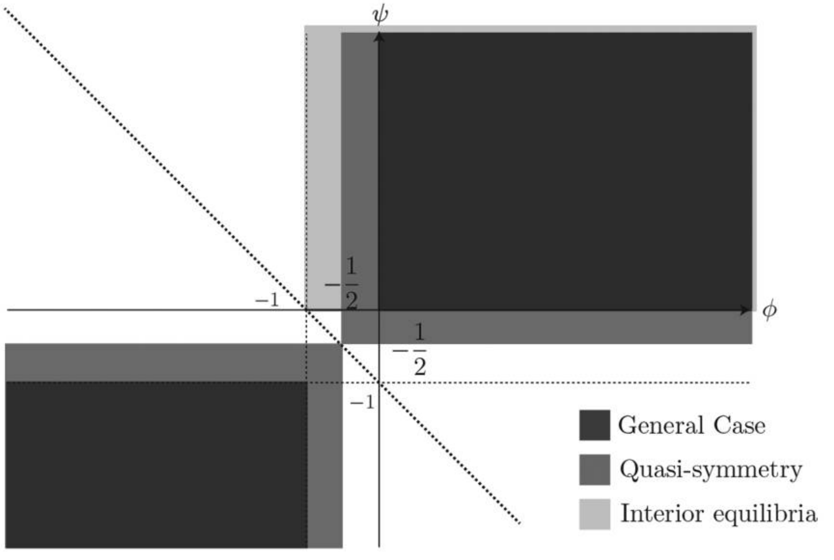

FIG. 1.—Existence and uniqueness: regions in (ϕ, ψ) space for which the gravity equilibrium is unique and interior. Existence can be guaranteed throughout the entire region except for the case $1 + \phi + \psi = 0$. A color version of this figure is available online.

does not exist in the knife-edge case where $\sigma = (1 + a)/(a + b)$, as agglomeration forces lead to the concentration of all economic activity in one location (see Allen and Arkolakis 2014).

As long as the partial elasticity of aggregate demand with respect to own output price is greater than -1 and the partial elasticity of supply with respect to the real output price is positive, all equilibria are interior. For example, in the economic geography model above, if these conditions are satisfied, one can show that the welfare of an uninhabited location approaches infinity as its population approaches zero, ensuring that all locations will be populated in equilibrium.

An equilibrium is unique as long as the partial elasticity of aggregate demand to output prices is negative (i.e., $\phi \geq 0$) and the partial elasticity of aggregate supply is positive (i.e., $\psi \geq 0$). There is also a unique interior equilibrium when the demand elasticity is positive and the supply elasticity is negative and both elasticities have magnitudes greater than one, although such parameter constellations are less economically meaningful (and there may also exist noninterior equilibria). Multiplicity of interior equilibria may arise in cases when supply and demand elasticities are both positive (which occurs, e.g., in trade models when goods are complements) or when supply and demand elasticities are both negative (which occurs, e.g., in economic geography models when agglomeration forces are stronger than dispersion forces). Such examples of multiplicity are easy to construct—appendix B.7

provides examples of multiplicity in a two-location world where either the demand elasticity is negative (in which case the relative demand and supply curves are both upward sloping) or the supply elasticity is negative (in which case the relative demand and supply curves are both downward sloping). Finally, quasi-symmetric trade frictions allow us to extend the range of gravity constants for which uniqueness is guaranteed, but do not qualitatively change the intuition for the results.

IV. The Network Effects of a Trade Shock

We now turn to how the universal gravity framework can be used to make predictions of how a change in trade frictions alter equilibrium trade flows, incomes, and real output prices in each location.[16]

To begin, we define two $N \times 1$ vectors (which, with some abuse of language, we call "curves"): the supply curve $\mathbf{Q}^s$, to be the set of supply equations (11) from C.3 (multiplied by output prices and divided by κ), and the demand curve $\mathbf{Q}^d$, to be the set of market clearing (demand) equations combining C.1, C.2, C.4, and C.5, that is,

$$\mathbf{Q}^s(\mathbf{p}, \mathbf{P}) \equiv \left[p_i \bar{c}_i \left(\frac{p_i}{P_i}\right)^{\psi}\right]_{i \in S}, \tag{14}$$

$$\mathbf{Q}^d(\mathbf{p}, \mathbf{P}, \Xi; \boldsymbol{\tau}) \equiv \left[\sum_{j \in S} \tau_{ij}^{-\phi} p_i^{-\phi} P_j^{\phi} p_j \bar{c}_j \left(\frac{p_j}{P_j}\right)^{\psi} \Xi \xi_j\right]_{i \in S}, \tag{15}$$

where $\mathbf{p} \equiv (p_i)_{i \in S}$ and $\mathbf{P} \equiv [(\Sigma_{j \in S} \tau_{ji}^{-\phi} p_j^{-\phi})^{-(1/\phi)}]_{i \in S}$ are $N \times 1$ vectors and $\tau \equiv (\tau_{ij})_{i,j \in S}$ is an $N^2 \times 1$ vector.[17] Note that we express both the supply and demand curves in value terms, which will prove helpful in deriving the comparative statics in terms of observed trade flows.

In equilibrium, supply is equal to demand, that is, $\mathbf{Q}^s(\mathbf{p}, \mathbf{P}) = \mathbf{Q}^d(\mathbf{p}, \mathbf{P}; \boldsymbol{\tau})$, and equation (5) is expressed as follows:

$$\sum_{i \in S} \mathbf{Q}_i^s(\mathbf{p}, \mathbf{P}) = \sum_{i \in S} \mathbf{Q}_i^d(\mathbf{p}, \mathbf{P}, \Xi; \boldsymbol{\tau}) \Leftrightarrow \sum_{i \in S} (1 - \Xi \xi_i) p_i \bar{c}_i \left(\frac{p_i}{P_i}\right)^{\psi} = 0.$$

For notational convenience, define $Z(\mathbf{p}, \mathbf{P}, \Xi)$ as $\Sigma_{i \in S}(1 - \Xi \xi_j) p_i C_i (p_i/P_i)^{\psi}$. We fully differentiate these equations, along with the definition of the price index, to yield the following system of $2N + 1$ linear equations relating a small change in trade costs, $D \ln \boldsymbol{\tau}$, to a small change in output prices and price indices, $D \ln \mathbf{p}$ and $D \ln \mathbf{P}$, respectively:

[16] In what follows, we focus on the policy shocks that alter bilateral trade frictions $\{\tau_{ij}\}_{i,j \in S}$. In app. B.8, we show how one can apply similar tools to characterize the theoretical properties and conduct counterfactuals in an Armington trade model with tariffs.

[17] One can also conduct comparative statics with respect to ξ. See Dekle, Eaton, and Kortum (2008).

$$\left(\underbrace{\begin{pmatrix} D_{\ln \mathbf{p}}\mathbf{Q}^s & \mathbf{0} & 0 \\ \mathbf{0} & \mathbf{I} & 0 \\ \mathbf{0} & \mathbf{0} & D_{\ln \Xi} Z \end{pmatrix}}_{\equiv \mathbf{S}} - \underbrace{\begin{pmatrix} D_{\ln \mathbf{p}}\mathbf{Q}^d & D_{\ln \mathbf{P}}\mathbf{Q}^d - D_{\ln \mathbf{P}}\mathbf{Q}^s & D_{\ln \Xi}\mathbf{Q}^d \\ D_{\ln \mathbf{p}} \ln \mathbf{P} & \mathbf{0} & 0 \\ -D_{\ln \mathbf{p}} Z & -D_{\ln \mathbf{P}} Z & 0 \end{pmatrix}}_{\equiv \mathbf{D}}\right)\begin{pmatrix} D \ln \mathbf{p} \\ D \ln \mathbf{P} \\ D \ln \Xi \end{pmatrix}$$
$$= \underbrace{\begin{pmatrix} D_{\ln \tau}\mathbf{Q}^d \\ D_{\ln \tau} \ln \mathbf{P} \\ \mathbf{0} \end{pmatrix}}_{\equiv \mathbf{T}} D \ln \boldsymbol{\tau},$$

where **S** (the *supply matrix*) and **D** (the *demand matrix*) are $2N + 1 \times 2N + 1$ matrices capturing the marginal effects of a change in the output price on the supply and demand curves (where the demand matrix also captures the net effect of a change in the price index), respectively, and **T** is a $2N + 1 \times N^2$ matrix capturing the marginal effects of a change in trade costs on the demand curve and price index.

Given expressions (14) and (15), we can write all three matrices solely as a function of the gravity constants and observables as follows:

$$\mathbf{S} = \begin{pmatrix} (1+\psi)\mathbf{Y} & \mathbf{0} & 0 \\ \mathbf{0} & \mathbf{I} & 0 \\ \mathbf{0} & \mathbf{0} & -\sum_{i \in S} E_i \end{pmatrix},$$
$$\mathbf{D} = \begin{pmatrix} -\phi\mathbf{E} + (1+\psi)\mathbf{X}, & (\phi - \psi)\mathbf{X} + \psi\mathbf{E}, & (E_i)_i \\ \mathbf{E}^{-1}\mathbf{X}^T & \mathbf{0} & 0 \\ (1+\psi)(Y_i - E_i)_i^T & -\psi(Y_i - E_i)_i^T & 0 \end{pmatrix}, \tag{16}$$

and

$$\mathbf{T} = \begin{pmatrix} -\phi(\mathbf{X} \otimes \mathbf{1}) \circ (\mathbf{I} \otimes \mathbf{1}) \\ \left(\mathbf{E}^{-1}\mathbf{X}^T \otimes \mathbf{1}\right) \circ (\mathbf{1} \otimes \mathbf{I}) \\ \mathbf{0} \end{pmatrix}, \tag{17}$$

where **X** is the (observed) $N \times N$ trade flow matrix whose $\langle i, j \rangle$th element is X_{ij}, **Y** is the $N \times N$ diagonal income matrix whose ith diagonal element is Y_i, **E** is the $N \times N$ diagonal income matrix whose ith diagonal element is E_i, **I** is the $N \times N$ identity matrix, **1** is an $1 \times N$ matrix of ones, $\mathbf{I}_i$ is the standard ith basis for $\mathbb{R}^N$, $\otimes$ represents the Kronecker product, and $\circ$ represents the element-wise multiplication (i.e., Hadamard product).[18]

A simple application of the implicit-function theorem allows us to characterize the elasticity of prices and price indices to any trade cost shock. Define the $2N+1 \times 2N+1$ matrix $\mathbf{A} \equiv \mathbf{S} - \mathbf{D}$, and, with a slight abuse of notation, let $A^{-1}_{k,l}$ denote the $\langle k, l \rangle$th element of the (pseudo)inverse of $\mathbf{A}$. Then we have the following theorem:

Theorem 2. Consider any model contained in the universal gravity framework. Suppose that $\mathbf{X}$ satisfies strong connectivity. If $\mathbf{A}$ has rank $2N$, then

i. the elasticities of output prices and output price indices are given by

$$\begin{aligned} \frac{\partial \ln p_l}{\partial \ln \tau_{ij}} &= -\phi X_{ij} A^{-1}_{l,i} + \frac{X_{ij}}{E_j} A^{-1}_{l,N+j} \quad \text{and} \\ \frac{\partial \ln P_l}{\partial \ln \tau_{ij}} &= -\phi X_{ij} A^{-1}_{N+l,i} + \frac{X_{ij}}{E_j} A^{-1}_{N+l,N+j}; \end{aligned} \tag{18}$$

ii. if the largest absolute value of eigenvalues of $\mathbf{S}^{-1}\mathbf{D}$ is less than one, then $\mathbf{A}^{-1}$ has the following series expansion:

$$\mathbf{A}^{-1} = \sum_{k=0}^{\infty} \left(\mathbf{S}^{-1}\mathbf{D}\right)^k \mathbf{S}^{-1};$$

and

iii. if trade frictions are quasi-symmetric, trade is balanced, and $\phi, \psi \geq 0$, then for all $i, l \in S$ and $j \neq i, l$,

$$\frac{\partial \ln(p_i/P_i)}{\partial \ln \tau_{il}}, \frac{\partial \ln(p_l/P_l)}{\partial \ln \tau_{li}} < \frac{\partial \ln\left(p_j/P_j\right)}{\partial \ln \tau_{il}}$$

$$\frac{\partial \ln(p_i Q_i/P_i)}{\partial \ln \tau_{il}}, \frac{\partial \ln(p_l Q_l/P_l)}{\partial \ln \tau_{li}} < \frac{\partial \ln\left(p_j Q_j/P_j\right)}{\partial \ln \tau_{il}},$$

and the inequalities have the opposite sign ($>$) if $\phi, \psi \leq -1$.

Proof. See appendix A2.

Recall from section III that knowledge of the output prices and price indices up to scale is sufficient to recover real output prices and—along with the normalization C.6—is sufficient to recover equilibrium trade flows, expenditures, and incomes.[19] As a result, part i of theorem 2 states that, given

[18] In what follows (apart from pt. iii of theorem 2), we do not assume that C.5 holds in the data, i.e., that income is necessarily equal to expenditure. Rather, we allow for income and expenditure to differ by a location-specific scalar; i.e., we allow for (exogenous) deficits.

[19] Because of homogeneity of degree 0, we can without loss of generality normalize one price; moreover, from Walras's law, if $2N$ equilibrium conditions hold, then the last equilibrium

gravity constants and observed data, the (local) counterfactuals of these variables for all models contained in the universal gravity framework are the same.[20]

The second part of theorem 2 provides a simple interpretation of the counterfactuals as a shock propagating through the trade network. Consider a shock that decreases the trade cost between i and j by a small amount $\partial \ln \tau_{ij}$, and define $(\mathbf{S}^{-1}\mathbf{D})^k\mathbf{S}^{-1}$ as the k*th-degree effect* of the shock. It turns out the kth-degree effect is simply the effect of the $(k-1)$th-degree shock on the output prices and price indices of all locations' trading partners, holding constant their trading partners' prices and price indices. To see this, consider first the zeroth-degree effect. Holding constant the prices and price indices in all other locations, the direct effect of a decrease in $\partial \ln \tau_{ij}$ is a shift of the demand curve upward in i by $\phi X_{ij} \times \partial \ln \tau_{ij}$ and a decrease in the price index in j by $(X_{ij}/E_j) \times \partial \ln \tau_{ij}$. To reequilibriate supply and demand (holding constant prices and price indices in all other locations), we then trace along the supply curve to where supply equals demand by scaling the effect by $\mathbf{S}^{-1}$, for a total effect of $\mathbf{S}^{-1}\partial \ln \tau$. Consider now the first-degree effect. We first take the resulting changes in the price and the price index from the zeroth-degree effect and calculate how they shift the demand curve (and alter the price index) in all i and j trading partners by multiplying the zeroth-degree effect by the demand matrix, that is, $\mathbf{D}(\mathbf{S}^{-1}\partial \ln \tau)$. To find how this changes the price and the price index in each trading partner (holding constant the prices and price indices in the trading partners' trading partners), we then trace along the supply curve by again scaling the shock by $\mathbf{S}^{-1}$, for a combined effect of $\mathbf{S}^{-1}\mathbf{D}\mathbf{S}^{-1}\partial \ln \tau$. The process continues iteratively, with the kth-degree effect shifting the demand curve and price index according to the $(k-1)$th-degree shock and then reequilibrating supply and demand by tracing along the supply curve (holding constant the prices and price indices in all trading partners), for

condition holds as well. As a result, $\mathbf{A}$ will have at most $2N$ rank, and $\mathbf{A}^{-1}$ can be calculated by simply eliminating one row and column of $\mathbf{A}$ and then calculating its inverse. The values of the eliminated row can then be determined by using the normalization C.6. For example, if one removes the first row and column, $\partial \ln p_1/\partial \ln \tau_{ij}$ can be chosen to ensure that $\Sigma_{i\in S}(\partial \ln Y_i/\partial \ln \tau_{ij}) = 0$, so that C.6 is satisfied.

[20] In app. B.9, we show how the "exact hat algebra" (Dekle, Eaton, and Kortum 2008; Costinot and Rodríguez-Clare 2013) can be applied to any model in the universal gravity framework to calculate the effect of any (possibly large) trade shock. The key takeaway—that counterfactual predictions depend only on observed data and the value of the gravity constants—remains true globally. However, if the uniqueness conditions of theorem 1 do not hold, we are unaware of any procedure that guarantees that the solution found using the "exact hat algebra" approach corresponds to the counterfactual of the observed equilibrium. Indeed, it is straightforward to construct a simple example where, in the presence of multiple equilibria, iterative algorithms used to solve the "exact hat algebra" system of equations will converge to equilibria qualitatively different from what is observed in the data even for arbitrarily small shocks, implying arbitrarily large counterfactual elasticities. In contrast, the elasticities in theorem 2 will provide the correct local counterfactual elasticities around the observed equilibrium even in the presence of multiple equilibria.

an effect of $(\mathbf{S}^{-1}\mathbf{D})^k\mathbf{S}^{-1}\partial \ln \tau$, as claimed.[21] The total change in prices and price indices is the infinite sum of all kth-degree shocks.

The third part of theorem 2 says that the direct impact of a symmetric decline in trade frictions $\partial \ln \tau_{il}$ and $\partial \ln \tau_{li}$ on real output prices (and real expenditure) in the directly affected locations i and l will be larger than the impact of that shock in any other indirectly affected location $j \neq i, l$. If the demand and supply elasticities are positive, then a decline in trade frictions will cause the real output prices in the directly affected locations to rise more than in any indirectly affected location (the ordering is reversed if the demand and supply elasticities are negative). This analytical result characterizes the relative impact of a trade friction shock on different locations in a model with many locations and arbitrary bilateral frictions.[22]

V. Estimating the Gravity Constants

In the previous section, we saw that the impact of a trade friction shock on trade flows, incomes, expenditures, and real output prices in any gravity model can be determined solely from observed trade flow data and the value the demand and supply elasticities. In this section, we show how these gravity constants can be estimated. We use data on international trade flows, so for the remainder of the paper we refer to a location as a country.

A. *Methodology*

We first derive an equation that shows that the relationship between three observables—relative trade shares, relative incomes, and relative own expenditure shares—are governed by the two gravity constants. We then show how this relationship, under minor assumptions, can be used as an estimating equation to recover the gravity constants. We begin by combining C.1 and C.2 to express the expenditure share of country j on trade from i, relative to its expenditure on its own goods as a function of the trade frictions, the output prices in i and j, and the aggregate-demand elasticity:

$$\frac{X_{ij}}{X_{jj}} = \left(\frac{\tau_{jj}p_j}{\tau_{ij}p_i}\right)^{\phi}.$$

We then use the relationship $p_i = Y_i/Q_i$ to rewrite this expression in terms of incomes and aggregate quantities and rely on C.3 to write the

[21] One can also derive the alternative representation $\mathbf{A}^{-1} = -\Sigma_{k=0}^{\infty}\mathbf{D}^{-1}(\mathbf{S}\mathbf{D}^{-1})^k$, in which the ordering is reversed: the kth-degree effect is calculated by first shifting the supply curve by the $(k-1)$th-degree shock and then tracing along the demand curve to reequilibriate supply and demand.

[22] Mossay and Tabuchi (2015) prove a similar result in a three-country world.

equilibrium output as a function of output prices and the output price index:

$$\frac{X_{ij}}{X_{jj}} = \left[\frac{\tau_{jj}\left(Y_j/\bar{c}_j\right)\left(p_i/P_i\right)^{\psi}}{\tau_{ij}\left(Y_i/\bar{c}_i\right)\left(p_j/P_j\right)^{\psi}}\right]^{\phi}. \tag{19}$$

We now define $\lambda_{jj} \equiv X_{jj}/E_j$ to be the fraction of income country j spends on its own goods (j's "own expenditure share"). By combining C.1 and C.2, we note that j's own expenditure share can be written as $\lambda_{jj} = [\tau_{jj}(p_j/P_j)]^{-\phi}$, which allows us to write equation (19) (in log form) as

$$\ln\frac{X_{ij}}{X_{jj}} = -\phi\ln\frac{\tau_{ij}}{\tau_{jj}} + \phi\ln\frac{Y_j}{Y_i} + \psi\ln\frac{\lambda_{jj}}{\lambda_{ii}} - \phi\ln\frac{\bar{c}_j}{\bar{c}_i} + \phi\psi\ln\frac{\tau_{jj}}{\tau_{ii}}. \tag{20}$$

Equation (20) shows that the demand elasticity ϕ is equal to the partial elasticity of trade flows to relative incomes, whereas the supply elasticity ψ is equal to the partial elasticity of trade flows to the relative own expenditure shares. Intuitively, the greater j's income relative to i (holding all else equal, especially the relative supply shifters $\ln(\bar{c}_j/\bar{c}_i)$), the greater the price in j relative to i, and hence the more it would demand from i relative to j; the greater the demand elasticity ϕ, the greater the effect of the price difference on expenditure. Conversely, because the real output price is inversely related to a country's own expenditure share, the greater j's own expenditure share relative to i, the lower the relative aggregate supply to j and hence the more j will consume from i relative to j; the larger the supply elasticity ψ, the more responsive supply will be to differences in own expenditure share.

Equation (20) forms the basis of our strategy for estimating the gravity elasticities ϕ and ψ. However, it also highlights two important challenges in estimation. First, because unobserved trade frictions act as a residual in equation (20), we require a moment condition along with observed trade flows in order to estimate the gravity elasticities.[23] Second, equation (20) highlights that the gravity elasticities are partial elasticities holding the (unobserved) relative supply shifters $\{\bar{c}_i\}_{i\in S}$ fixed. Because both income and own expenditure shares are correlated with supply shifters through the equilibrium structure of the model, any estimation procedure must contend with this correlation between observables and unobservables.

[23] Relatedly, app. B.10 shows how for any observed set of trade flows $\{X_{ij}\}$ and any assumed set of gravity elasticities $\{\phi, \psi\}$, own trade frictions $\{\tau_{ii}\}$, and supply shifters $\{\bar{c}_i\}$, there will exist a unique set of trade frictions $\{\tau_{ij}\}_{i\neq j}$ for which the observed trade flows are the equilibrium trade flows of the model.

In order to address both concerns, we combine plausibly exogenous observed geographic variation with the general equilibrium structure of the model to estimate the gravity elasticities. We proceed in a two-stage procedure.[24] First, we rewrite equation (20) as

$$\ln \frac{X_{ij}}{X_{jj}} = -\phi \ln \frac{\tau_{ij}}{\tau_{jj}} - \ln \pi_i + \ln \pi_j,$$

where $\ln \pi_i \equiv \phi \ln Y_i + \psi \ln \lambda_{ii} - \phi \ln \bar{c}_i + \phi\psi \ln \tau_{ii}$ is a country-specific fixed effect. We assume that relative trade frictions scaled by the trade elasticity can be written as a function of their continent of origin c, continent of destination d, and the decile of distance between the origin and destination countries, l:

$$-\phi \ln \frac{\tau_{ij}}{\tau_{jj}} = \beta^l_{cd} + \varepsilon_{ij},$$

where ε_{ij} is a residual assumed to be independent across origin-destination pairs. The country-specific fixed effect can then be recovered from the following equation:

$$\ln \frac{X_{ij}}{X_{jj}} = \beta^l_{cd} - \ln\pi_i + \ln\pi_j + \varepsilon_{ij}, \tag{21}$$

where we estimate β^l_{cd} nonparametrically, using a set of 360 dummy variables (10 distance deciles $\times$ 6 origin continents $\times$ 6 destination continents). Let $\ln\hat{\pi}_i$ denote the estimated fixed effect, and define $\hat{\nu}_i \equiv \ln\hat{\pi}_i - \ln\pi_i$ to be its estimation error.

In the second stage, we write the estimated fixed effect as a function of income and own expenditure share:

$$\ln\hat{\pi}_i = \phi \ln Y_i + \psi \ln\lambda_{ii} + \nu_i, \tag{22}$$

[24] While the two-step procedure we follow resembles the procedure used in Eaton and Kortum (2002) to recover the trade elasticity from observed wages, there are two important differences. First, our procedure applies to a large class of trade and economic geography models and allows us to simultaneously estimate both the demand (trade) elasticity and the supply elasticity (rather than assuming, e.g., that the population of a country is exogenous and calibrating the model to a particular intermediate good share). Second, our procedure relies on the general equilibrium structure of the model to generate the identifying variation (rather than, e.g., instrumenting for wages with the local labor supply, which would be inappropriate for economic geography models).

where $\nu_i \equiv -\phi \ln \bar{c}_i + \phi\psi \ln \tau_{ii} + \hat{\nu}_i$ is a residual that combines the unobserved supply shifter, the unobserved own trade friction, and the estimation error from the first stage. As mentioned above, it is not appropriate to estimate equation (22) via ordinary least squares (OLS), as variation in the supply shifter will affect income and the own expenditure share through the equilibrium structure of the model, creating a correlation between the residual and the observed covariates. Intuitively, the larger the supply shifter of a country, the greater its output and hence the greater the trade flows for a given observed income; since the country-specific fixed effect $\ln\pi_i$ is decreasing in relative trade flows, the OLS estimate of ϕ will be biased downward.

To overcome this bias, we pursue an instrumental variables (IV) strategy, where we use the general equilibrium structure of the model to construct a valid instrument. To do so, we calculate the equilibrium trade flows of a hypothetical world where the bilateral trade frictions and supply shifters depend only on observables. We then use the incomes and relative own expenditure shares of this hypothetical world as instruments for the observed incomes and own expenditure shares. These "model-implied" instruments are valid as long as (1) they are correlated with their observed counterparts (which we can verify) and (2) the observable components of the bilateral trade frictions and supply shifters are uncorrelated with unobserved supply shifters.

Because the first-stage estimation of equation (21) provides an unbiased estimate of $-\phi \ln(\tau_{ij}/\tau_{jj})$, we use the estimated origin-continent-destination-continent-decile coefficients $\hat{\beta}^l_{cd}$ to create our counterfactual measure of bilateral trade frictions (normalizing own trade frictions $\tau_{jj} = 1$). In the simplest version of our procedure, we then calculate the equilibrium income and own expenditure share, given these bilateral trade frictions, assuming that the supply shifter $\bar{c}_i$ is equal in all countries. In this version of the procedure, the instrument is valid as long as the general equilibrium effects of distance on the origin fixed effects of a gravity equation are uncorrelated with unobserved heterogeneity in supply shifters (or own trade frictions). Because we calculate the equilibrium of the model in a counterfactual world where there is no heterogeneity in supply shifters, it seems reasonable to assume that the resulting equilibrium income and own expenditure shares that we use as instruments are uncorrelated with any real-world heterogeneity. However, our instrument would be invalid if there were a correlation between unobserved supply shifters and the observed geography of a country (e.g., if countries more remotely located were also less productive or less attractive places to reside).

To mitigate such a concern (and to allow for more realistic variation across countries in supply), we extend the approach to allow the supply shifter to vary across countries, depending on a vector of (exogenous) observables X^c_i, for example, land controls such as the amount of fertile land, geographic controls such as the distance to the nearest coast, institutional controls such as the rule of law, historical controls such as the population in 1400, and schooling and R&D controls such as average years of schooling. Given a set of supply shifters $\{\bar{c}_i\}$ that depend only on these observables and the set of trade frictions that depend only on our nonparametric estimates from above, we recalculate the equilibrium income and own expenditure share in each country. We then use the equilibrium values from this hypothetical world as our instruments, while controlling directly for the observables X_i^c in equation (22). As a result, the identifying variation from the instruments arises only through the general equilibrium structure of the model.[25] Intuitively, differences in observables such as land area in neighboring countries generates variation in the demand that a country faces for its production as well as variation in the price it faces for its consumption, even conditional on its own observables.

There are two things to note about the above procedure. First, to construct the hypothetical equilibrium incomes and own expenditure shares requires assuming values of the gravity constants ϕ and ψ for the hypothetical world. In what follows, we choose a demand elasticity $\phi = 8.28$ and a supply elasticity $\psi = 3.76$, which correspond to the (estimated) demand elasticity and (implicitly calibrated) supply elasticity in Eaton and Kortum (2002). We should note that while the particular choice of the these parameters will affect the strength of the constructed instruments, they will not affect the consistency of our estimates of the gravity constants under the maintained assumption that bilateral distances are uncorrelated with the unobserved supply shifters, conditional on observables.[26]

The second thing to note about the estimation procedure is more subtle. As mentioned in section III and discussed in detail in appendix B.3, when bilateral trade frictions are "quasi-symmetric" the equilibrium origin and destination shifters will be equal up to scale. In this case, there will be a perfect log-linear relationship between the income of a country, its own expenditure share and its supply shifter.[27] As a result, if we were to impose

[25] Calculating the counterfactual equilibrium income and own expenditure share in each country when the supply shifters depend on observables requires assuming a particular mapping between the observables X_i^c and the supply shifter $\bar{c}_i$. We assume that $\bar{c}_i = X_i^c\beta^c$ and note that the theory implies the following equilibrium condition:

$$\ln Y_i = \frac{\phi}{\phi - \psi}\ln \bar{c}_i + \frac{1+\psi}{\psi - \phi}\ln \gamma_i + \frac{\psi}{\psi - \phi}\ln \delta_i.$$

As a result, we choose the β^c that arise from the OLS regression $\ln Y_i = [\phi/(\phi - \psi)]X_i^c\beta^c + \epsilon_i$. Although our estimates of β^c may be biased because of the correlation between X_i^c and ϵ_i, this bias affects only the strength of the instrument, because if each X_i^c is uncorrelated with the residual ν_i in eq. (22) (i.e., X_i^c is exogenous), then any linear combination of X_i^c will also be uncorrelated with the residual.

[26] In principle, we could search over different values of the gravity constants to find the constellation that maximizes the power of our instruments. In practice, however, our estimates vary only a small amount across different values of the gravity constants.

[27] In particular, $(1 + 2\phi)\ln E_i = (2\phi)\ln \bar{c}_i + (1 - 2\psi)\ln \lambda_{ii} + C$.

quasi-symmetric bilateral trade frictions in the hypothetical world, the equilibrium income and expenditure shares generated would be perfectly collinear, preventing us from simultaneously identifying the demand and supply elasticities in the second stage. Intuitively, identification of the demand elasticity requires variation in a country's supply curve (its destination fixed effect), whereas identification of the supply elasticity requires variation in a country's demand curve (its origin fixed effect); when trade frictions are quasi-symmetric, however, the two covary perfectly. Our choice to allow distance to affect trade frictions differently, depending on the continent of origin and continent of destination, introduces the necessary asymmetries in the trade frictions to allow the model-implied instruments to vary separately, allowing for identification of both the supply and demand elasticities simultaneously. To address concerns about the extent to which these asymmetries are sufficient to separately identify the two, we report the Sanderson-Windmeijer F-test (see Sanderson and Windmeijer 2016) in the results that follow.

B. *Data*

We now briefly describe the data we use to estimate the gravity constants.

Our trade data come from the Global Trade Analysis Project (GTAP), version 7 (Narayanan and Walmsley 2008). These data provide bilateral trade flows between 94 countries for the year 2004. To construct own trade flows, we subtract total exports from the total sales of domestic product, that is, $X_{ii} = X_i - \Sigma_{j \neq i} X_{ij}$. We use the bilateral distances between countries from the CEPII (Centre d'études prospectives et d'informations internationales) gravity data set of Head, Mayer, and Ries (2010) to construct deciles of distance between two countries. We rely on the data set of Nunn and Puga (2012) to provide a number of country-level characteristics that plausibly affect supply shifters, including "land controls" (land area interacted with the fraction of fertile soil, desert, and tropical areas), "geographic controls" (distance to the nearest coast and the fraction of country within 100 km of an ice-free coast), "historical controls" (log population in 1400 and the percentage of the population of European descent), and "institutional controls" (the quality of the rule of law). Finally, following Eaton and Kortum (2002), we also consider "schooling and R&D controls" including the average years of schooling from UNESCO (2015) and the R&D stocks from Coe, Helpman, and Hoffmaister (2009), where a dummy variable is included if the country is not in each respective data set.

C. *Estimation Results*

Table 2 presents the results of our estimation of equation (20). Column 1 presents the OLS regression; we estimate a positive supply elasticity and a

TABLE 2
ESTIMATING THE GRAVITY CONSTANTS

	OLS	IV					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log income (demand elasticity)	−.403**	1.484	3.278	4.364*	3.882**	3.539***	3.715***
	(.171)	(1.157)	(2.674)	(2.371)	(1.838)	(1.356)	(1.312)
Log own expenditure share (supply elasticity)	3.381**	92.889***	108.592**	116.649**	71.859**	64.968**	68.488**
	(1.600)	(13.417)	(48.104)	(47.944)	(35.883)	(33.014)	(32.198)
Land controls	No	No	Yes	Yes	Yes	Yes	Yes
Geographic controls	No	No	No	Yes	Yes	Yes	Yes
Historical controls	No	No	No	No	Yes	Yes	Yes
Institutional controls	No	No	No	No	No	Yes	Yes
Schooling and R&D controls	No	No	No	No	No	No	Yes
First-stage Sanderson-Windmeijer *F*-test:							
Income		25.909	3.994	6.349	20.095	34.198	25.763
p-value		.004	.102	.053	.007	.002	.004
Own expenditure share		72.702	4.388	4.923	3.561	4.577	5.205
p-value		.000	.090	.077	.118	.085	.071
Observations	94	94	94	94	94	94	94

NOTE.—The dependent variable is the estimated country fixed effect of a gravity regression of the log ratio of bilateral trade flows to destination own trade flows on categorical deciles of distance variables, where the coefficient is allowed to vary by continent of origin and destination. Hence, each observation in the regressions above is a country. Instruments for income and own expenditure share are the equilibrium values from a trade model where the bilateral trade frictions are those predicted from the same gravity equation and countries are either identical in their supply shifters (col. 2) or their supply shifters are estimated from a regression of observed income on observables (cols. 3–7). In the latter case, the observables determining the supply shifters are controlled for directly in the second-stage regression, so identification of the demand and supply elasticities arise only from the general equilibrium effect of distance on income and own expenditure shares. Land controls include land area interacted with fractions fertile soil, desert, and tropical areas. Geographic controls include the distance to the nearest coast and the fraction of country within 100 km of an ice-free coast. Historical controls include the log population in 1400 and the percentage of the population of European descent. Institutional controls include the quality of the rule of law. Schooling and R&D controls are average years of schooling (from UNESCO) and the R&D stocks (from Coe, Helpman, and Hoffmaister 2009), where a dummy variable is included if the country is not in each respective data set. Land, geographic, and historical control are from Nunn and Puga (2012). Standard errors clustered at the continent level are reported in parentheses.

* $p < .10$.
** $p < .05$.
*** $p < .01$.

negative demand elasticity, consistent with the discussion above that the OLS estimate of the demand elasticity is biased downward. Column 2 presents the IV estimation, where the counterfactual income and own expenditure shares comprising our instrument are constructed assuming equal supply shifters. After correcting for the bias arising from the correlation between the unobserved supply shifters and observed incomes and own expenditure shares, we find positive supply and demand elasticities, although the demand elasticity is not statistically significant. Columns 3–7 sequentially allow the supply shifter in the construction of the instrument to vary across countries, depending on an increasing number of observables (while including these same observables as controls in both the first and second stages of the IV estimation of eq. [20]). Including these observables both increases the strength of the instruments and reduces the concern that the instruments are correlated with unobserved supply shifters. Reassuringly, our estimated demand and supply elasticities vary only slightly with the inclusion of additional controls.[28]

In our preferred specification (col. 7), we estimate a demand elasticity of $\phi = 3.72$ (95% confidence interval: [1.14,6.29]) and a supply elasticity of $\psi = 68.49$ (95% confidence interval: [5.38,131.60]).[29] Hence, our demand elasticity estimate is somewhat lower than the preferred estimate, in Eaton and Kortum (2002), of 8.28 (although similar to their estimate of 3.6, using variation in wages), as well as similar to estimates of trade elasticity of around 4 in Anderson and van Wincoop (2004), Simonovska and Waugh (2014), and Donaldson (2018). Unlike these papers, however, here we also estimate the supply elasticity. Our point estimate, while noisily estimated, is substantially larger than and statistically different (at the 5% level) from the supply elasticity to which Eaton and Kortum (2002) implicitly calibrate. Moreover, our estimated value is consistent with recent estimates of labor mobility from the migration literature. To see this, consider an economic geography framework with intermediate goods, agglomeration forces, and Fréchet-distributed preferences over location (see the last row of table 1). If we match the labor share in production of 0.21 in Eaton and Kortum (2002) and the agglomeration force of $\alpha = 0.10$ in Rosenthal and Strange (2004), then our point estimate of ψ is consistent with a migration elasticity (Fréchet shape parameter) of 1.4. This is similar to

[28] Figure 4 (figs. 3–12 are available in app. B) shows that our IVs of counterfactual income and own expenditure shares are positively correlated with their observed counterparts, even after differencing out the observables in the supply shifters.

[29] While the p-value of the Sanderson-Windmeijer F-test is statistically significant in the first stage for income, it is only marginally statistically significant for expenditure shares, suggesting that the wide confidence interval for the supply elasticity may be due in part to a weak instrument.

estimates from the migration literature using observed labor flows and about one-third to one-half the size of within-country estimates.[30]

VI. The Impact of a United States–China Trade War

We now apply the estimates from section V to evaluate the impact of a trade war between the United States and China. We model the trade war as an increase in the trade frictions between the United States and China (holding constant all other trade frictions). We then characterize how such a trade war propagates through the trade network, using the methodology developed in section IV.[31]

There are two zeroth-degree effects of the trade war: first, the United States and China export less to each other, causing the output prices in both countries to fall; second, the cost of importing increases, causing the price index in both countries to rise. Both effects cause the real output price to decline, with a greater decline in China because both its export and import shares with the United States are relatively larger.

Figure 2*A* depicts the first-degree effect on the real output price in all countries. The effect in the United States and China is positive, as the zeroth-degree decline in output price reduces the cost of own expenditure (causing the price index to fall in both countries). In other countries, however, the first-degree effect is negative, as the United States and China demand less of their goods, causing their trading partner's output prices to fall. The most negatively affected countries are those that export the most to the United States and China.

Summing across all degree shocks yields the total elasticity of real output prices in each country to the trade war shock, which figure 2*B* depicts.[32] Not surprisingly, the two countries hurt most by a trade war are the United States and China. Moreover, while all countries are made worse off, the countries

[30] Ortega and Peri (2013) estimate a migration elasticity to destination country income of 0.6, using international migration flows and an estimate of 1.8 for the subsample of migration flows within the European Union, albeit not using a log-linear gravity specification. Within countries (and with log-linear gravity specifications), Monte, Redding, and Rossi-Hansberg (2015) estimate a migration elasticity of 4.4 in the United States, Tombe and Zhu (2015) estimate a migration elasticity of 2.54 in China, and Morten and Oliveira (2014) estimate a migration elasticity of 3.4 in Brazil.

[31] In the counterfactuals that follow, we accommodate the deficits observed in the data by assuming that the observed ratio of expenditure to income for each country remains constant and impose an aggregate market-clearing condition that total income is equal to total expenditure. The results are qualitatively similar if we instead solve for the (unique) set of balanced trade flows that match the observed import shares and treat these balanced trade flows as the data.

[32] Figures 5–8 depict, respectively, the impact of degrees 0, 1, 2, and higher (and fig. 9 the total impact) on the relative prices, relative output, income, the relative price index, and real output prices in each country.

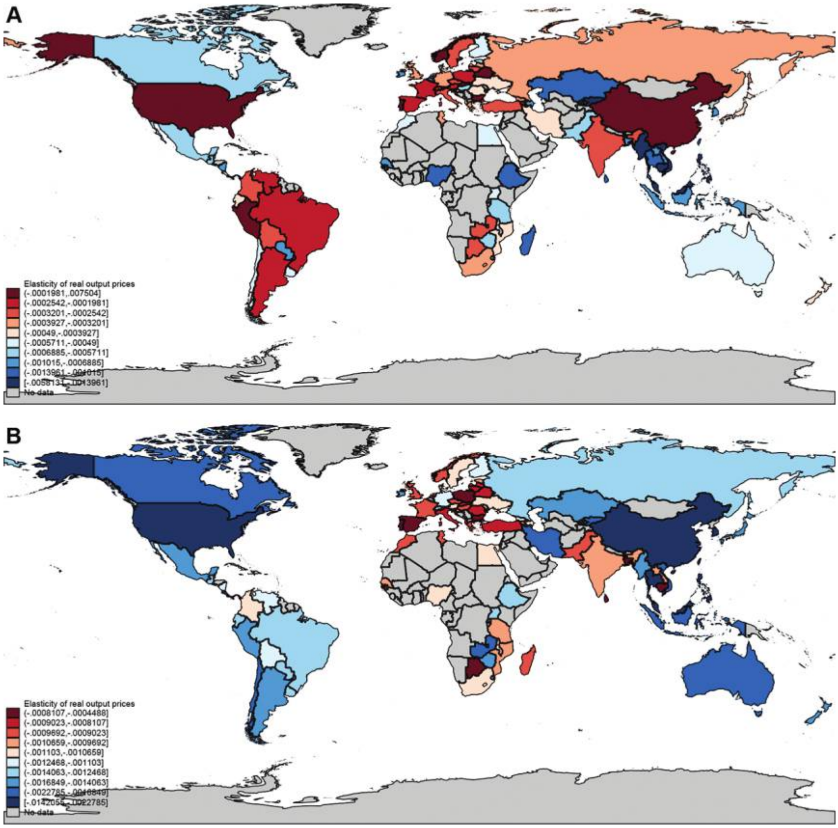

FIG. 2.—Network effect of a United States–China trade war. This figure depicts the elasticity of real output prices to an increase in the bilateral trade frictions between the United States and China (a "trade war") in all countries. *A*, The first-degree effect, which is the effect of the direct shock to the United States and China on all countries through the trade network, holding the output prices and quantities of their trading partners fixed. *B*, Total effect of the trade war on the real output price in each country.

that are closely linked through the trading network with the United States and China (e.g., Canada, Mexico, Vietnam, and Japan) are hurt more than those countries that are less connected (e.g., India). All told, we estimate that a 10% increase in bilateral trade frictions is associated with a decline in real output price of 0.04% in the United States and 0.14% in China. These modest changes in the real output price are due to the large supply elasticity, causing the aggregate output to reallocate away from the United States and China in response to the trade war. The converse of this result, however, is that the reallocation of the aggregate output results in large changes to total real expenditure: for example, in the Armington trade model interpretation, a 10% increase in bilateral trade frictions causes the total real expenditure to fall by 2.7% in the United States and by 9.8% in China.[33]

There are two potential concerns about these estimated effects. First, because the elasticities correspond to an infinitesimal shock, one may worry that the effects of a large trade war may differ. To address this concern, we calculate the effect of a 50% increase in bilateral trade frictions, using the methodology discussed in appendix B.9. The correlation between the local elasticities and global changes exceeds 0.99, indicating that the local relative effect of the trade war is virtually the same as the global effect (see fig. 10). However, the local effect does overstate the global effect of such a shock, as we find that log first differences implied by the global shock are roughly 80% the size of those implied by the local elasticities. Second, the effects of the trade war above were calculated, given the gravity constants estimated in section V; one may be concerned that the effects of the trade wars may differ substantially across alternative values of these elasticities. To address this concern, we calculate the effects of a trade war for a large number of different combinations of supply and demand elasticities (see fig. 11). Across all constellations in the 95% confidence interval of the two estimated gravity constants, the calculated elasticities are quite similar, with a 10% increase in bilateral trade frictions associated with a decline in real output price between 0.03% and 0.05% in the United States and between 0.07% and 0.26% in China. Of course, as section IV emphasizes, the particular value of the gravity constants may substantially affect the impact of counterfactuals more generally.

VII. Conclusion

In this paper, we provide a framework that unifies a large set of trade and geography models. We show that the properties of models within this framework depend crucially on the value of two gravity constants: the aggregate-supply and aggregate-demand elasticities. Sufficient conditions for the existence and uniqueness of the equilibria depend solely on the gravity constants. Moreover, given observed trade flows, these gravity constants are sufficient to determine the effect of a trade friction shock on trade flows, incomes, and real output price without needing to specify a particular underlying model.

We then develop a new model-implied IV approach for estimating the gravity constants, using the general equilibrium structure of the framework. Using our estimates, we find that potentially large losses may arise as a result of a trade war between the United States and China occur.

[33] Recall from sec. II that while the changes in real output prices are identified from the value of trade flows alone, without specifying κ in eq. (11), the change in total real expenditure is identified only up to scale. In Armington trade models with intermediates, however, this is not a problem, as $\kappa = 1$.

By providing a universal framework for understanding the general equilibrium forces in trade and geography models, we hope that this paper provides a step toward unifying the quantitative general equilibrium approach with the gravity regression analysis common in the empirical trade and geography literature. Toward this end, we have developed a tool kit that operationalizes all the theoretical results presented in this paper.[34] We also hope that the tools developed here can be extended to understand other general equilibrium spatial systems, such as those incorporating additional types of spatial linkages beyond trade frictions.

Appendix A

Proofs

A1. Proof of Theorem 1

A1.1. Part i

The proof proceeds as follows. First, we transform the equilibrium conditions to the associated nonlinear integral equations form. However, we cannot directly apply the fixed-point theorem for the nonlinear integral equations, since the system does not map to a compact space. Therefore, we need to "scale" the system so that we can apply the fixed point, which implies that there exists a fixed point for the scaled system. Finally, we construct a fixed point for the original nonlinear integral equations. In this subsection, we show how to set up in the associated integral equation form and apply the fixed-point theorem. The other technical parts are proven in appendix B.4. Note that our result proposition is a natural generalization of Karlin and Nirenberg (1967) to a system of nonlinear integral equations.

Define z as follows:

$$z \equiv \begin{pmatrix} (x_i)_i \\ (y_i)_i \end{pmatrix} \equiv \begin{pmatrix} \left(p_i^{1+\psi+\phi} P_i^{-\psi}\right)_i \\ \left(P_i^{-\phi}\right)_i \end{pmatrix}.$$

Then the system of equations (6) and (7) of the general equilibrium gravity model is rewritten in vector form:

$$\begin{pmatrix} (x_i)_i \\ (y_i)_i \end{pmatrix} = \begin{pmatrix} \sum_j K_{ij} \bar{c}_i^{-1} \bar{c}_j x_j^{a_{11}} y_j^{a_{12}} \\ \sum_j K_{ji} x_j^{a_{21}} y_j^{a_{22}} \end{pmatrix}, \tag{23}$$

where $A = (a_{ij})_{i,j}$ is given by

$$A = \begin{pmatrix} \dfrac{1+\psi}{1+\psi+\phi} & -\dfrac{1+\phi}{1+\psi+\phi} \\ -\dfrac{\phi}{1+\psi+\phi} & \dfrac{\psi}{1+\psi+\phi} \end{pmatrix}.$$

[34] The tool kit is available for download on Allen's website.

Also the kernel, K_{ij}, is given by $K_{ij} = \tau_{ij}^{-\phi}$. Note that we cannot directly apply Browser's fixed-point theorem for equation (23), since there is no trivial compact domain for that equation. Therefore, consider the following "scaled" version of equation (23):

$$z = \begin{pmatrix} (x_i)_i \\ (y_i)_i \end{pmatrix} = \begin{pmatrix} \dfrac{\sum_j K_{ij}\bar{c}_i^{-1}\bar{c}_j x_j^{a_{11}} y_j^{a_{12}}}{\sum_{i,j} K_{ij}\bar{c}_i^{-1}\bar{c}_j x_j^{a_{11}} y_j^{a_{12}}} \\ \dfrac{\sum_j K_{ji} x_j^{a_{21}} y_j^{a_{22}}}{\sum_{i,j} K_{ji} x_j^{a_{21}} y_j^{a_{22}}} \end{pmatrix} \equiv F(z), \tag{24}$$

and F is defined over the following compact set C:

$$C = \{x \in \Delta(R_+^N); x_i \in [\underline{x}, \bar{x}] \ \forall\ i\} \times \left\{y \in \Delta(R_+^N); y_i \in [\underline{y}, \bar{y}] \ \forall\ i\right\}, \tag{25}$$

where the bounds for x and y are given as follows:

$$\bar{x} \equiv \max_{i,j} \frac{K_{ij}\bar{c}_i^{-1}\bar{c}_j}{\sum_{i,j} K_{ij}\bar{c}_i^{-1}\bar{c}_j}, \quad \underline{x} \equiv \min_{i,j} \frac{K_{ij}\bar{c}_i^{-1}\bar{c}_j}{\sum_{i,j} K_{ij}\bar{c}_i^{-1}\bar{c}_j},$$

$$\bar{y} \equiv \max_{i,j} \frac{K_{ji}}{\sum_{i,j} K_{ji}}, \qquad \underline{y} = \min_{i,j} \frac{K_{ji}}{\sum_{i,j} K_{ji}}.$$

It is trivial to show that F maps from C to C and continuous over the compact set C, so that we can apply Brouwer's fixed point and there exists an fixed point $z^* \in C$.

There are two technical points to be proved: first, there exists a fixed point for the original (unscaled) system (eq. [23]); and second, the equilibrium z^* is strictly positive. These two claims are proved in lemmas 1 and 2, respectively, in appendix B.4.

A1.2. Part iii

It suffices to show that there exists a unique interior solution for equation (23). Suppose that there are two strictly positive solutions (x_i, y_i) and $(\hat{x}_i, \hat{y}_i)$ such that there does not exist $t, s > 0$ satisfying

$$(x_i, y_i) = (t\hat{x}_i, s\hat{y}_i).$$

Namely, the two solutions are "linearly independent." First note that for any $i \in S$, we can evaluate the first row of equation (23).

$$\frac{x_i}{\hat{x}_i} = \frac{1}{\hat{x}_i} \sum_{j \in S} K_{ij}\bar{c}_i^{-1}\bar{c}_j \left(\frac{x_j}{\hat{x}_j}\right)^{\alpha_{11}} \left(\frac{y_j}{\hat{y}_j}\right)^{\alpha_{12}} (\hat{x}_j)^{\alpha_{11}} (\hat{y}_j)^{\alpha_{12}} \tag{26}$$

$$\leqslant \max_{j \in S} \left(\frac{x_j}{\hat{x}_j}\right)^{\alpha_{11}} \max_{j \in S} \left(\frac{y_j}{\hat{y}_j}\right)^{\alpha_{12}}. \tag{27}$$

Taking the maximum of the left-hand side,

$$\max_{i\in S}\frac{x_i}{\hat{x}_i} \leqslant \max_{j\in S}\left(\frac{x_j}{\hat{x}_j}\right)^{\alpha_{11}} \max_{j\in S}\left(\frac{y_j}{\hat{y}_j}\right)^{\alpha_{12}}. \tag{28}$$

Lemma 3, in appendix B.4, shows that the inequality is actually strict. Analogously, we obtain

$$\min_{i\in S}\frac{x_i}{\hat{x}_i} \geqslant \min_{j\in S}\left(\frac{x_j}{\hat{x}_j}\right)^{\alpha_{11}} \min_{j\in S}\left(\frac{y_j}{\hat{y}_j}\right)^{\alpha_{12}}. \tag{29}$$

Dividing equation (28) by equation (29) shows that

$$1 \leqslant \mu_x \equiv \frac{\max_{i\in S}(x_i/\hat{x}_i)}{\min_{i\in S}(x_i/\hat{x}_i)} < \frac{\max_{j\in S}\left(x_j/\hat{x}_j\right)^{\alpha_{11}}}{\min_{j\in S}\left(x_j/\hat{x}_j\right)^{\alpha_{11}}} \times \frac{\max_{j\in S}\left(y_j/\hat{y}_j\right)^{\alpha_{12}}}{\min_{j\in S}\left(y_j/\hat{y}_j\right)^{\alpha_{12}}} = \mu_x^{|\alpha_{11}|} \times \mu_y^{|\alpha_{12}|},$$

where

$$\mu_y \equiv \frac{\max_{i\in S}(y_i/\hat{y}_i)}{\min_{i\in S}(y_i/\hat{y}_i)}.$$

The same argument is applied to the second row of equation (23) to obtain the following inequality:

$$1 \leqslant \mu_y \equiv \frac{\max_{i\in S}(y_i/\hat{y}_i)}{\min_{i\in S}(y_i/\hat{y}_i)} < \frac{\max_{j\in S}\left(x_j/\hat{x}_j\right)^{\alpha_{21}}}{\min_{j\in S}\left(x_j/\hat{x}_j\right)^{\alpha_{21}}} \times \frac{\max_{j\in S}\left(y_j/\hat{y}_j\right)^{\alpha_{22}}}{\min_{j\in S}\left(y_j/\hat{y}_j\right)^{\alpha_{22}}} = \mu_x^{|\alpha_{21}|} \times \mu_y^{|\alpha_{22}|}.$$

Taking logs in the two inequalities and exploiting the restriction, we can write

$$\begin{pmatrix} \ln \mu_x \\ \ln \mu_y \end{pmatrix} < \underbrace{\begin{pmatrix} |\alpha_{11}| & |\alpha_{12}| \\ |\alpha_{21}| & |\alpha_{22}| \end{pmatrix}}_{=|A|} \begin{pmatrix} \ln \mu_x \\ \ln \mu_y \end{pmatrix}, \tag{30}$$

which from the Collatz-Wielandt formula implies that the largest eigenvalue of $|A|$ is greater than one. However, we prove in lemma 4, in appendix B.4, that the sufficient condition in part ii of theorem 1 guarantees that the largest absolute eigenvalue is 1. As a result, this is a contradiction.

A1.3. Quasi Symmetry

When the bilateral trade frictions satisfy quasi symmetry, then we can reduce the system to an N-dimensional integral system (see app. B.3). Then the same logic used above can be applied to show there exists a unique strictly positive solution. As mentioned above, this result follows directly from Karlin and Nirenberg (1967) and is summarized in theorem 2.19 of Zabreyko et al. (1975). QED

A2. Proof of Theorem 2

A2.1. Part i

Equation (18) is a direct application of the implicit-function theorem. Define a function $F: R^{2N} \to R^{2N}$ as follows:

$$F_i\left((\ln p_i)_{i=1}^N, (\ln P_i)_{i=1}^N\right) = \kappa \bar{c}_i p_i^{1+\psi} P_i^{-\psi} - \kappa \sum_k \tau_{ik}^{-\phi} p_i^{-\phi} \bar{c}_k P_k^{\phi-\psi} p_k^{1+\psi} \Xi \xi_k,$$

$$F_{N-1+i}\left((\ln p_i)_{i=1}^N, (\ln P_i)_{i=1}^N\right) = P_i^{-\phi} - \sum_k \tau_{ki}^{-\phi} p_k^{-\phi}.$$

Applying the implicit function theorem for F, we obtain the comparative statics (eq. [18]).

A2.2. Part ii

Note that $\mathbf{A}$ is written as follows:

$$\mathbf{A} = \mathbf{S}\left(\mathbf{I} - \mathbf{S}^{-1}\mathbf{D}\right),$$

where $\mathbf{S}$ and $\mathbf{D}$ are defined by equation (16). If the largest absolute eigenvalue for $\mathbf{S}^{-1}\mathbf{D}$ is less than one, then $\mathbf{A}^{-1}$ is expressed as $\Sigma_{k=0}^{\infty}(\mathbf{S}^{-1}\mathbf{D})^k\mathbf{S}^{-1}$. Note that we could have similarly written $\mathbf{A} = -(\mathbf{I} - \mathbf{S}\mathbf{D}^{-1})\mathbf{D}$, so that if the largest eigenvalue for $\mathbf{S}\mathbf{D}^{-1}$ is less than one, $\mathbf{A}^{-1}$ can be expressed as $-\Sigma_{k=0}^{\infty}\mathbf{D}^{-1}(\mathbf{S}\mathbf{D}^{-1})^k$, as noted in note 22.

A2.3. Part iii

When quasi-symmetric assumption and balanced trade are imposed, destination effects are proportional to the associated origin effects. Therefore, as shown in appendix B.3, the equilibrium is characterized by the following single nonlinear system of equations:

$$\underbrace{p_i^{1+\psi-\psi[(1+\psi+\phi)/(\psi-\phi)]}\left(\frac{\tau_i^A}{\tau_i^B}\right)^{-\psi[\phi/(\psi-\phi)]}(\bar{c}_i)^{[\phi/(\psi-\phi)]}}_{=Y_i/\kappa} = \sum_{j\in S}\underbrace{\tilde{\tau}_{ij}^{-\phi} p_i^{-\phi}\left(\tau_i^A\right)^{-\phi}\left(\tau_j^A\right)^{-\phi} p_j^{-\phi}}_{=X_{ij}/\kappa}. \tag{31}$$

As above, define z_i for all $i \in S$ as follows:

$$z_i(p;\tau) = \kappa p_i^{1+\psi-\psi[(1+\psi+\phi)/(\psi-\phi)]}\left(\frac{\tau_i^A}{\tau_i^B}\right)^{-\psi[\phi/(\psi-\phi)]}(\bar{c}_i)^{[\phi/(\psi-\phi)]} - \kappa\sum_{j\in S}\tilde{\tau}_{ij}^{-\phi} p_i^{-\phi}\left(\tau_i^A\right)^{-\phi}\left(\tau_j^A\right)^{-\phi} p_j^{-\phi}.$$

Then apply the implicit-function theorem to equation (31),

$$\frac{\partial \ln p}{\partial \ln \tau_{il}} = -2\underbrace{\left(\frac{\partial z}{\partial \ln p}\right)^{-1}}_{N\times N}\underbrace{\frac{\partial z}{\partial \ln \tau_{il}}}_{N\times 1}. \tag{32}$$

Note that the number 2 shows up to preserve the quasi symmetry of trade frictions. As in the general trade friction case, $\partial z/\partial \ln p$ is expressed as observables:

$$\frac{\partial z}{\partial \ln p} = \left(\phi\frac{1+\psi+\phi}{\phi-\psi}\right)\left(\mathbf{Y}+\frac{\phi-\psi}{1+\psi+\phi}\mathbf{X}\right),$$

where $\mathbf{Y} = \operatorname{diag}(Y_i)$ and $\mathbf{X} = (X_{ij})_{i,j\in S}$. Define $\mathbf{A}$ as follows:

$$\mathbf{A} = \mathbf{Y} + \frac{\phi-\psi}{1+\psi+\phi}\mathbf{X}.$$

From lemma 5, $\mathbf{A}$ has positive diagonal elements and is dominant of its rows. Equation (32) is

$$\frac{\partial \ln p_i}{\partial \ln \tau_{il}} = -2\frac{\phi-\psi}{1+\psi+\phi}A_{ii}^{-1}X_{il},$$
$$\frac{\partial \ln p_j}{\partial \ln \tau_{il}} = -2\frac{\phi-\psi}{1+\psi+\phi}A_{ji}^{-1}X_{il}.$$

Since the price index is log-linear with respect to the associated output price, we have

$$\frac{\partial \ln P_i}{\partial \ln \tau_{il}} = \frac{1+\psi+\phi}{\psi-\phi}\frac{\partial \ln p_i}{\partial \ln \tau_{il}}.$$

Therefore, the real output price is

$$\frac{\partial \ln(p_i/P_i)}{\partial \ln \tau_{il}} = \left(\frac{2\phi+1}{\phi-\psi}\right)\frac{\partial \ln p_i}{\partial \ln \tau_{il}} = -2\frac{2\phi+1}{1+\psi+\phi}A_{ii}^{-1}X_{il}.$$

Then the ordering of the real output price follows from part iii of theorem 2, $A_{ii}^{-1} > A_{ji}^{-1}$ for $j \in S - i$. The result for real expenditure then follows immediately from C.5 and equation (11), as $E_i/P_i \propto \bar{c}_i(p_i/P_i)^{1+\psi}$:

$$\frac{\partial \ln(p_i Q_i/P_i)}{\partial \ln \tau_{il}} = -2\frac{2\phi+1}{1+\psi+\phi}(1+\psi)A_{ii}^{-1}X_{il} + \underbrace{\frac{\partial \ln \kappa}{\partial \ln \tau_{il}}}_{\text{common}}.$$

By the same argument, the ordering of $\partial \ln(p_i Q_i/P_i)/\partial \ln \tau_{il}$ follows. QED

References

Adao, Rodrigo, Arnaud Costinot, and Dave Donaldson. 2017. "Nonparametric Counterfactual Predictions in Neoclassical Models of International Trade." *A.E.R.* 107 (3): 633–89.

Allen, David (Treb). 2012. "Information Frictions in Trade." PhD diss., Yale Univ. http://search-4.library.yale.edu/catalog/11288714.

Allen, Treb, and Costas Arkolakis. 2014. "Trade and the Topography of the Spatial Economy." *Q.J.E.* 129 (3): 1085–140.

Allen, Treb, Costas Arkolakis, and Xiangliang Li. 2014. "On the Existence and Uniqueness of Trade Equilibria." Manuscript, Northwestern Univ. and Yale Univ.

Alvarez, Fernando, and Robert E. Lucas Jr. 2007. "General Equilibrium Analysis of the Eaton-Kortum Model of International Trade." *J. Monetary Econ.* 54 (6): 1726–68.

Anderson, James E. 1979. "A Theoretical Foundation for the Gravity Equation." *A.E.R.* 69 (1): 106–16.

Anderson, James E., Mario Larch, and Yoto V. Yotov. 2015. "Growth and Trade with Frictions: A Structural Estimation Framework." Working Paper no. 21377 (July), NBER, Cambridge, MA.

Anderson, James E., and Eric van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *A.E.R.* 93 (1): 170–92.

———. 2004. "Trade Costs." *J. Econ. Literature* 42 (3): 691–751.

Anderson, James E., and Yoto V. Yotov. 2010. "The Changing Incidence of Geography." *A.E.R.* 100 (5): 2157–86.

Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare. 2019. "The Elusive Pro-Competitive Effects of Trade." *Rev. Econ. Studies* 86 (1): 46–80.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare. 2012. "New Trade Models, Same Old Gains?" *A.E.R.* 102 (1): 94–130.

Arkolakis, Costas, Svetlana Demidova, Peter J. Klenow, and Andrés Rodríguez-Clare. 2008. "Endogenous Variety and the Gains from Trade." *A.E.R.* 98 (2): 444–50.

Armington, P. S. 1969. "A Theory of Demand for Products Distinguished by Place of Production." *Staff Papers (Internat. Monetary Fund)* 16:159–78.

Arrow, Kenneth J., and Frank Hahn. 1971. *General Competitive Analysis.* Amsterdam: North-Holland.

Baldwin, Richard, and Daria Taglioni. 2006. "Gravity for Dummies and Dummies for Gravity Equations." Working Paper no. 12516 (September), NBER, Cambridge, MA.

Bartelme, Dominick. 2014. "Trade Costs and Economic Geography: Evidence from the US." Manuscript.

Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum. 2003. "Plants and Productivity in International Trade." *A.E.R.* 93 (4): 1268–90.

Berry, Steven, Amit Gandhi, and Philip Haile. 2013. "Connected Substitutes and Invertibility of Demand." *Econometrica* 81 (5): 2087–111.

Bosker, Maarten, and Bastian Westbrook. 2016. "A Theory of Trade in a Global Production Network." Working Paper no. 14-14, Utrecht School Econ.

Caliendo, Lorenzo, and Fernando Parro. 2015. "Estimates of the Trade and Welfare Effects of NAFTA." *Rev. Econ. Studies* 82 (1): 1–44.

Chaney, Thomas. 2008. "Distorted Gravity: The Intensive and Extensive Margins of International Trade." *A.E.R.* 98 (4): 1707–21.

Cheng, Leonard. 1985. "Inverting Systems of Demand Functions." *J. Econ. Theory* 37 (1): 202–10.

Coe, David T., Elhanan Helpman, and Alexander W. Hoffmaister. 2009. "International R&D Spillovers and Institutions." *European Econ. Rev.* 53 (7): 723–41.

Costinot, Arnaud, Dave Donaldson, and Ivana Komunjer. 2012. "What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas." *Rev. Econ. Studies.* 79 (2): 581–608.

Costinot, Arnaud, and Andrés Rodríguez-Clare. 2013. "Trade Theory with Numbers: Quantifying the Consequences of Globalization." Working Paper no. 18896 (March), NBER, Cambridge, MA.

Costinot, Arnaud, Andrés Rodríguez-Clare, and Iván Werning. 2016. "Micro to Macro: Optimal Trade Policy with Firm Heterogeneity." Working Paper no. 21989 (February), NBER, Cambridge, MA.

Dekle, Robert, Jonathan Eaton, and Samuel Kortum. 2008. "Global Rebalancing with Gravity: Measuring the Burden of Adjustment." *IMF Staff Papers* 55 (3): 511–40.

Di Giovanni, Julian, and Andrei A. Levchenko. 2010. "Putting the Parts Together: Trade, Vertical Linkages, and Business Cycle Comovement." *American Econ. J.: Macroeconomics* 2 (2): 95–124.

———. 2013. "Firm Entry, Trade, and Welfare in Zipf's World." *J. Internat. Econ.* 89 (2): 283–96.

Donaldson, Dave. 2018. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." *A.E.R.* 108 (4–5): 899–934.

Donaldson, Dave, and Richard Hornbeck. 2012. "Railroads and American Economic Growth: New Data and Theory." Manuscript.

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography and Trade." *Econometrica* 70 (5): 1741–79.

Egger, Peter H., and Serey Nigai. 2015. "Structural Gravity with Dummies Only: Constrained ANOVA-Type Estimation of Gravity Models." *J. Internat. Econ.* 97 (1) 86–99.

Evmorfopoulos, Nestor. 2012. "Some Results on Diagonally Dominant Matrices with Positive Diagonal Elements." Tech. report, Univ. Thessaly.

Fajgelbaum, Pablo D., and Amit K. Khandelwal. 2014. "Measuring the Unequal Gains from Trade." Working Paper no. 20331 (July), NBER, Cambridge, MA.

Fally, Thibault. 2015. "Structural Gravity and Fixed Effects." *J. Internat. Econ.* 97 (1): 76–85.

Frankel, Jeffrey A., and David H. Romer. 1999. "Does Trade Cause Growth?" *A.E.R.* 89 (3): 379–99.

Head, Keith, and Thierry Mayer. 2013. "Gravity Equations: Workhorse, Toolkit, and Cookbook." In *Handbook of International Economics*, vol. 4, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 131–195. Amsterdam: North-Holland.

Head, Keith, Thierry Mayer, and John Ries. 2010. "The Erosion of Colonial Trade Linkages after Independence." *J. Internat. Econ.* 81 (1): 1–14.

Head, Keith, Thierry Mayer, and Mathias Thoenig. 2014. "Welfare and Trade without Pareto." *A.E.R.* 104 (5): 310–16.

Helpman, Elhanan. 1998. "The Size of Regions." In *Topics in Public Economics: Theoretical and Applied Analysis*, edited by David Pines, Efraim Sadka, and Itzhak Zilcha, 33–54. Cambridge: Cambridge Univ. Press.

Johnson, Charles R., and Ronald L. Smith. 2011. "Inverse *M*-Matrices, II." *Linear Algebra and Its Applications* 435 (5): 953–83.

Karlin, Samuel, and Lewis Nirenberg. 1967. "On a Theorem of P. Nowosad." *J. Math. Analysis and Applications* 17 (1): 61–67.

Krugman, Paul. 1980. "Scale Economies, Product Differentiation, and the Pattern of Trade." *A.E.R.* 70 (5): 950–59.

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green. 1995. *Microeconomic Theory*. Oxford: Oxford Univ. Press.

McKenzie, Lionel W. 1959. "On the Existence of General Equilibrium for a Competitive Market." *Econometrica* 27 (1): 54–71.

———. 1961. "On the Existence of General Equilibrium: Some Corrections." *Econometrica* 29 (2): 247–48.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6): 1695–725.

Melitz, Marc J., and Stephen Redding. 2015. "New Trade Models, New Welfare Implications." *A.E.R.* 105 (3): 1105–46.

Menezes, António Gomes de, Mário Fortuna, Francisco Silva, and José Cabral Vieira. 2006. "Computable General Equilibrium Models: A Literature Review." Working Paper no. 06/2006. Centro Estudos Econ. Aplicada Atlântico, Ponta Delgada, Portugal.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg. 2015. "Commuting, Migration and Local Employment Elasticities." Working Paper no. 21706 (November), NBER, Cambridge, MA.

Morten, Melanie, and Jacqueline Oliveira. 2014. "Migration, Roads and Labor Market Integration: Evidence from a Planned Capital City." Manuscript.

Mossay, Pascal, and Takatoshi Tabuchi. 2015. "Preferential Trade Agreements Harm Third Countries." *Econ. J.* 125 (589): 1964–85.

Narayanan, Badri G., and Terri L. Walmsley, eds. 2008. *Global Trade, Assistance, and Production: The GTAP 7 Data Base.* West Lafayette, IN: Center Global Trade Analysis, Purdue Univ.

Novy, Dennis. 2010. "International Trade Without CES: Estimating Translog Gravity." CEP Discussion Paper no. 1031, Centre Econ. Performance, London.

Nunn, Nathan, and Diego Puga. 2012. "Ruggedness: The Blessing of Bad Geography in Africa." *Rev. Econ. and Statis.* 94 (1): 20–36.

Ortega, Francesc, and Giovanni Peri. 2013. "The Effect of Income and Immigration Policies on International Migration." *Migration Studies* 1 (1): 47–74.

Polyanin, Andrei D., and Alexander V. Manzhirov. 2008. *Handbook of Integral Equations*, 2nd edition. Boca Raton, FL: Chapman & Hall/CRC.

Redding, Stephen J. 2016. "Goods Trade, Factor Mobility and Welfare." *J. Internat. Econ.* 101:148–67.

Redding, Stephen J., and Esteban A. Rossi-Hansberg. 2017. "Quantitative Spatial Economics." *Ann. Rev. Econ.* 9:21–58.

Redding, Stephen J., and Daniel M. Sturm. 2008. "The Costs of Remoteness: Evidence from German Division and Reunification." *A.E.R.* 98 (5): 1766–97.

Rosenthal, Stuart S., and William C. Strange. 2004. "Evidence on the Nature and Sources of Agglomeration Economies." In *Handbook of Regional and Urban Economics*, vol. 4, edited by J. Vernon Henderson and Jacques-François Thisse, 2119–71. Amsterdam: North-Holland.

Sanderson, Eleanor, and Frank Windmeijer. 2016. "A Weak Instrument *F*-Test in Linear IV Models with Multiple Endogenous Variables." *J. Econometrics* 190 (2): 212–21.

Simonovska, Ina, and Michael E. Waugh. 2014. "The Elasticity of Trade: Estimates and Evidence." *J. Internat. Econ.* 92 (1): 34–50.

Tombe, Trevor, and Xiaodong Zhu. 2015. "Trade, Migration and Productivity: A Quantitative Analysis of China." Working Paper no. 542, Dept. Econ., Univ. Toronto.

UNESCO. 2015. "Welcome to UIS.Stat." Montreal: UNESCO Inst. Statis. http://data.uis.unesco.org/.

Waugh, Michael E. 2010. "International Trade and Income Differences." *A.E.R.* 100 (5): 2093–124.

Zabreyko, P. P., A. I. Koshelev, M. A. Krasnosel'skii, S. G. Mikhlin, L. S. Rakovshchik, and V. Y. Stetsenko. 1975. *Integral Equations: A Reference Text.* Leyden: Noordhoff Internat.