

Generalized Separability and the Gains from Trade

Thibault Fally*
UC Berkeley ARE

September 2018

Abstract

The vast majority of models in international trade make strong assumptions on the structure of preferences, either assuming homothetic, directly-separable or indirectly-separable preferences. These assumptions lead to sharp implications for the quantitative gains from trade, conditional on estimated elasticities of trade to trade costs and observed import penetration. This note examines the gains from trade under a more flexible demand system that encompasses the three cases of demand usually imposed, while retaining the property that prices can be summarized by a single aggregator. Combined with monopolistic competition and Pareto distributions of marginal costs on the supply side (which leads to gravity equations), this demand system yields more flexible responses of prices to income and trade, and a wider range of predictions for the gains from trade. To estimate the gains from trade, this note briefly describes an identification strategy exploiting price responses to changes in market access and trade costs.

Keywords: Gains from trade, Gravity, Generalized Separability, Single aggregator, Non-homothetic preferences.

JEL Classification: F10, D11, D43, L11

*UC Berkeley, Agricultural and Resource Economics Department, NBER and CEPR. Contact: fally@berkeley.edu. Many many thanks to Andrés Rodríguez-Clare for very helpful comments.

1 Introduction

How large are the gains from international trade? Following Arkolakis, Costinot and Rodríguez-Clare (2012), a large body of work has examined the gains from trade through the lens of various types of trade models, using import penetration as a sufficient statistic. The conclusions of Arkolakis et al. (2012) have been extended by examining alternative assumptions on both the demand side and the supply side.¹ Looking into the demand side, recent work has shown that it is crucial to depart from homothetic constant-elasticity (CES) preferences to allow for variable markups, and non-trivial price responses.² However, existing papers have focused on specific types of preferences. In particular, Arkolakis, Costinot, Donaldson and Rodríguez-Clare (2015, henceforth ACDR) focus on directly-separable as well as homothetic preferences, whereas Bertolotti, Etro and Simonovska (2018) focus on indirectly-separable preferences.

This paper unifies and generalizes these approaches by considering the generalized Gorman-Pollak demand system³ which encompasses CES, directly-separable, indirectly separable, and special cases of homothetic preferences.⁴ Such demand system allows for more flexible price and income effects than directly or indirectly-separable preferences, while retaining the property that prices can be summarized by a single aggregator (generalized separability), which is particularly convenient under monopolistic competition. This general form of demand has been seldom used—perhaps for the lack of formal statements on integrability conditions, which are now provided in a companion paper (Fally, 2018).⁵ On the supply side, the model incorporates standard assumptions of recent trade models (e.g. as in ACDR, Chaney, 2008, Melitz, 2003), monopolistic competition with heterogeneous firms and Pareto distributions of marginal costs.

With generalized Gorman-Pollak demand, the model generates a response of price levels to per-capita income, as in Bertolotti et al (2018), while retaining a dependence of prices to trade shocks as in ACDR. With variable markups and incomplete pass-through of costs to prices, prices do not only depend on marginal costs but also on reservation prices (choke prices) in the destination market, leading to pricing to market. Generalized separability allows for a more flexible relationship between the choke price, income levels, and the degree of competition in the destination market, which itself depend on trade costs and proximity to third markets. The gains from trade can still be expressed as a simple function of the changes in import penetration but the formula leads to a wider range of estimates, conditional on markup and trade elasticities. Indirectly-additive preferences provide a lower bound of the gains from trade and homothetic preferences provide a higher bound. Within these bounds, all intermediate outcomes are theoretically possible.

Rather than assuming a specific shape of preferences, these results indicate that more precise

¹See, e.g., Feenstra (2018a) and Costinot and Rodríguez-Clare (2018) for recent surveys.

²Models extending Melitz (2003) to incorporate variable markups include Arkolakis et al. (2015), Bertolotti et al. (2018), Simonovska (2015), Behrens et al. (2014), Feenstra (2018b), among others.

³See Gorman (1972, 1995), Pollak (1972), Fally (2018).

⁴This homothetic case corresponds to the homothetic single aggregator (HSA) demand system studied in Matsuyama and Ushchev (2017), which yields the same outcomes as the implicitly-additive homothetic preferences in ACDR.

⁵A recent exception is Bertolotti and Etro (2017).

estimates of demand systems are necessary to quantify the gains from trade. In the last section, I discuss an empirical strategy to estimate the gains from trade with such a demand system. I consider a semi-parameterization of demand that encompasses directly and indirectly-additive preferences, as well as homothetic demand with a single price aggregator. Assuming that econometricians also have estimates of trade and markup elasticities,⁶ the key additional parameter to estimate determines how choke prices vary with income and trade costs. To estimate this additional parameter, a suggestion is to use firm-level data on export prices for a given source country and regress these prices on changes in income and market access of destination countries (as measured for instance in Head and Mayer, 2010, and Redding and Venables, 2004).

2 A flexible demand system

This section describes the demand side, and the next section provides results from the full general-equilibrium model. The notation remains close to Arkolakis et al., 2015 (ACDR), which we can refer to for more details, especially on the supply side.

2.1 Demand system

Consumer income and total expenditures are denoted by w . For each good, we assume that demand is determined by its price p , consumer income w and an aggregator Λ :

$$q(p, w, \Lambda) = Q(\Lambda)D(F(\Lambda)p/w)$$

where $\Lambda = \Lambda(p, w)$ is itself a scalar function of all prices and income, homogeneous of degree zero in (w, p) . Λ is implicitly defined by the budget constraint, i.e. it is the implicit solution of:

$$\int_{\omega \in \Omega} p(\omega)Q(\Lambda)D(F(\Lambda)p(\omega)/w)d\omega = w \quad (1)$$

where Ω denotes the set of goods available to the consumer and $p(\omega)$ is the price of each variety $\omega \in \Omega$.

Conditions for integrability Assuming that all three functions D , F and Q are differentiable, and assuming that D is sufficiently downward sloping and elastic, i.e. $\varepsilon_D \equiv -\frac{d \log D}{d \log p} > 1$, the following two conditions are sufficient to ensure that such demand system is integrable (Fally 2018), i.e. can be derived from rational consumer behavior and utility maximization:

- i) $\varepsilon_F(\Lambda)\varepsilon_D(y) - \varepsilon_Q(\Lambda)$ has the same sign for all Λ and y
- ii) For any set of prices $p(\omega)$ and income w , equation (1) admits a solution in Λ .

where ε_Q and ε_F denote the elasticity of Q and F with respect to Λ .

⁶For instance, Simonovska and Waugh (2014) and Amiti et al. (2014) respectively provide estimates of the trade and markup elasticities that are consistent with the framework that we consider here.

Special cases These preferences are more general than typically assumed in the trade literature. They correspond to:

- Directly-separable preferences when $Q(\Lambda)$ is constant and $F(\Lambda) = \Lambda$.
- Indirectly-separable preferences when $F(\Lambda)$ is constant and $Q(\Lambda) = \Lambda$.
- Homothetic Single Aggregator (Matsuyama and Ushchev 2018) when $F(\Lambda) = Q(\Lambda) = \Lambda$.
- Iso-elastic shifters, a useful parameterization: $Q(\Lambda) = \alpha\Lambda^\beta$ and $F(\Lambda) = \Lambda^\gamma$.

In the case with iso-elastic shifters Q and F , it is without loss of generality to assume that $\gamma \in \{0, 1\}$ (where the case $\gamma = 0$ corresponds to indirectly-additive preferences).⁷ If $\gamma = 1$, a sufficient condition for integrability is that $\beta < \varepsilon_D$. Since we assume that $\varepsilon_D > 1$, a sufficient condition is $\beta < 1$. Note that β can be negative, which also ensures integrability. Indirectly-additive preferences can be seen as the limit case $\beta \rightarrow -\infty$, while directly-separable preferences correspond to $\beta = 0$.

2.2 Choke prices

Assume that $D(\cdot)$ is equal to zero when prices are high enough (i.e. above a reservation price, or choke price). Without loss of generality, assume that $D(x) = 0$ if and only if $x \geq 1$. Let us denote by P the choke price, the price above which demand is null. It is then equal to:

$$P = \frac{w}{F(\Lambda)} \quad (2)$$

A central observation by Simonovska (2015) and Bertoletti et al. (2018) is that prices are more strongly correlated across countries with per capita income than with market size or population. This motivates the use of indirectly-additive preferences rather than directly-additive preferences with which it is difficult to quantitatively reproduce the role of income. With indirectly-additive preferences, however, F is constant and the choke price is proportional to income. In contrast, the preferences used here allow for more flexibility in the general case.

To fully characterize how the choke price varies with income in the more general case, we need to account for how income affects Λ . Differentiating w.r.t w , we obtain:

$$\frac{d \log \Lambda}{d \log w} = \frac{\int_{\omega \in \Omega} \lambda(\omega) \varepsilon_D(p(\omega)/P) d\omega - 1}{\varepsilon_F \int_{\omega \in \Omega} \lambda(\omega) \varepsilon_D(p(\omega)/P) d\omega - \varepsilon_Q} \quad (3)$$

where $\lambda(\omega)$ denotes the share of variety ω in total expenditures. Hence:

$$\frac{d \log P}{d \log w} = 1 - \varepsilon_F \frac{d \log \Lambda}{d \log w} = \frac{\varepsilon_F - \varepsilon_Q}{\varepsilon_F \int_{\omega \in \Omega} \lambda(\omega) \varepsilon_D(p(\omega)/P) d\omega - \varepsilon_Q}$$

⁷If $\gamma \neq 0$, consider the change in variable $\Lambda' = \Lambda^\gamma$.

where ε_D is the elasticity of $1/D$. In the iso-elastic case with $F(\Lambda) = \Lambda$ and $Q(\Lambda) = \alpha\Lambda^\beta$, we have:

$$\frac{d \log P}{d \log w} = \frac{1 - \beta}{\int_{\omega \in \Omega} \lambda(\omega) \varepsilon_D(p(\omega)/P) d\omega - \beta}$$

where β can take any value between $-\infty$ and 1. As β is assumed to be smaller than both 1 and ε_D , richer consumers have higher choke prices and consume a larger variety of goods—an empirically-relevant property (see e.g. Hummels and Klenow, 2005).

2.3 Markups under monopolistic competition

Under monopolistic competition, we assume that each firm has a small market share and takes the price aggregator Λ as given. The price elasticity that a firm is facing is given the elasticity of function D , and the markup is equal to:

$$m = \frac{\varepsilon_D(p/P)}{\varepsilon_D(p/P) - 1} = \frac{\varepsilon_D(m c/P)}{\varepsilon_D(m c/P) - 1} \quad (4)$$

Conveniently, this is the same formula as in ACDR, because markups do not depend on Q and F once we know the choke price P . As described in Fally (2018), function D can be very flexible, leading to flexible patterns of markups across firms. As in ACDR, we further assume that the price elasticity ε_D is weakly increasing in the price p/P (a.k.a. “second law of demand”). This additional assumption ensures that equation (4) has a unique solution in markups for each P/c , and implicitly defines markups as a function of the choke-price-to-marginal-cost ratio:

$$m = \mu(P/c). \quad (5)$$

This assumption also implies that markups are larger for low-cost firms, which is in line with empirical evidence (see e.g. De Loecker and Warzynski, 2012).

3 Trade, prices and the gains from trade

We now embed these preferences in a trade model with monopolistic competition as in ACDR. We consider multiple asymmetric countries, indexed by i and j . There is a single factor, labor, that is inelastically supplied, with population L_i in each country i . Wages and per capita income are denoted by w_i , and L_i is also the number of consumers in country i . Demand is as described above, with a choke price denoted by $P_j = w_j/F(\Lambda_j)$ in each destination country j . We assume bilateral iceberg trade costs $\tau_{ij} \geq 1$ between i and j . This leads to a cost thresholds $c_{ij}^* = P_j/(w_i\tau_{ij})$. Firms in country i with marginal cost above $w_i c_{ij}^*$ face zero demand in country j if they also incur bilateral iceberg trade costs τ_{ij} .

We further assume that the marginal cost shifter c for each firm is drawn from a Pareto distribution

with coefficient $\theta > 1$, so that the cumulative distribution of the marginal cost is given by:

$$G_i(c) = b_i c^\theta$$

where b_i is a parameter describing average productivity in country i . Assuming that b_i is small enough, this implies that the distribution of marginal costs among firms in country i exporting to destination j is a Pareto distribution with coefficient θ and upper bound $w_i c_{ij}^*$. Finally, we assume for simplicity that the mass of firms N_i is fixed.⁸

3.1 Gravity

As in ACDR, we obtain a gravity equation thanks to the assumption that marginal costs are drawn from a Pareto distribution. This leads to aggregate trade between countries i and j equal to:

$$X_{ij} = \chi N_i b_i w_i^{-\theta} \times \tau_{ij}^{-\theta} \times L_j Q(\Lambda_j) P_j^{1+\theta} \quad (6)$$

where χ is a constant term (see Appendix). Trade X_{ij} is the product of a source-specific term $N_i b_i w_i^{-\theta}$, a dyadic trade cost term $\tau_{ij}^{-\theta}$, and importantly a destination-specific term $L_j Q(\Lambda_j) P_j^{1+\theta}$ which depends on both the destination choke price P_j and the demand shifter $Q(\Lambda_j)$. Note that the latter is constant when preferences are directly separable, as often assumed with gravity equations.

Let us remind that Λ_j is such that the budget constraint is satisfied, and that $P_j = w_j / F(\Lambda_j)$. Summing equation (6) for each destination j , this is equivalent to imposing:

$$\sum_i X_{ij} / L_j = \chi Q(\Lambda_j) (w_j / F(\Lambda_j))^{\theta+1} \left(\sum_i N_i b_i (w_i \tau_{ij})^{-\theta} \right) = w_j \quad (7)$$

which determines Λ_j , conditional on all wages.

With gravity equations and Pareto distributions, note also that aggregate profits are proportional to aggregate trade: $\Pi_{ij} = \zeta X_{ij}$ with $\zeta < 1$. As in ACDR, proportionality holds even if we allow for variable markups.

3.2 Effect of trade costs on prices

Let us normalize wages to unity $w_j = 1$ for a specific country j . Denote by ε_{Qj} and ε_{Fj} the elasticities of Q and F evaluated at Λ_j . By differentiating equation (7) we obtain:

$$d \log P_j = -\varepsilon_{Fj} d \log \Lambda_j = \frac{\theta \varepsilon_{Fj}}{(\theta + 1) \varepsilon_{Fj} - \varepsilon_{Qj}} \sum_i \lambda_{ij} d \log (w_i \tau_{ij}) \quad (8)$$

where λ_{ij} refers to share of country i in total import of country j . Using the insight of Arkolakis et al. (2012), this gravity framework with elasticity θ implies that the changes in import shares capture the changes in costs: $\sum_i \lambda_{ij} d \log (w_i \tau_{ij}) = \frac{d \log \lambda_{jj}}{\theta}$, and we obtain the following proposition:

⁸Note that the mass of firms remains constant under free entry with sunk costs and Pareto-distributed productivity.

Proposition 1 *For small changes in trade costs, the change in the choke price in country j is given by:*

$$d \log P_j = \frac{\varepsilon_{Fj}}{(\theta + 1)\varepsilon_{Fj} - \varepsilon_{Qj}} d \log \lambda_{jj} \quad (9)$$

Conditional on changes in import penetration $d \log \lambda_{jj}$ and the trade elasticity θ , the effect on price levels can take any value between 0 (indirectly-additive) and $-\frac{1}{\theta} d \log \lambda_{jj}$ (homothetic).

In the iso-elastic case with $F(\Lambda) = \Lambda$ and $Q(\Lambda) = \alpha \Lambda^\beta$ we have:

$$d \log P_j = \frac{1}{\theta + 1 - \beta} d \log \lambda_{jj} \quad (10)$$

We obtain the same formula as in ACDR but now with a continuum of β 's instead of imposing β to be equal to either 0 or 1. This semi-parameterization captures more simply the added flexibility that we obtain thanks to generalized separability, and is valid as long as β remains smaller than the elasticity of D (assumed to be larger than unity). It includes all tractable cases studied recently and provides a micro-foundation for intermediate cases. The homothetic case ($\beta = 1$) yields the largest effect on prices unless we assume a higher lower bound on price elasticities ε_D . Relative to homothetic preferences, the effect on prices is mitigated by a factor $\frac{\theta}{1+\theta}$ with directly-separable preferences ($\beta = 0$). The other extreme is obtained with indirectly-separable preferences (limit case $\beta = -\infty$) where changes in trade costs do not affect the choke price: $d \log P_j = 0$ (Bertoletti et al 2018).

Distribution of prices and markups In this framework, note that the choke price serves as a sufficient statistic for prices. Relative to the choke price, the distribution of prices remains invariant to changes in trade costs, given that the shape of the Pareto distribution is invariant to its truncation. Hence, Proposition 1 also describes how trade costs shift the entire distribution of prices.

In particular, as in ACDR, the distribution of markups remains invariant when marginal costs are drawn from Pareto. Markups are determined by the ratio of cost to the choke price and the function $\mu(P/c)$. This function does not vary across countries, and the distribution of P/c remains also invariant (Pareto distribution bounded at unity). This also implies that the average markup elasticity weighted by sales is also constant across countries and fully determined by functions μ , D and the Pareto parameter θ as in ACDR:

$$\rho = \int_1^\infty \frac{d \log \mu}{d \log v} \frac{\mu(v)/v D(\mu(v)/v) v^{-\theta-1} dv}{\int_1^\infty \mu(v')/v' D(\mu(v')/v') (v')^{-\theta-1} dv'} \quad (11)$$

One can notice that the distribution of markups (under monopolistic competition) does not depend on $Q(\Lambda)$ and $F(\Lambda)$, the demand and price shifters.

3.3 Gains from trade

We can now examine the effect of trade costs on welfare, combining the effect on price levels, markups and income. If we maintain the same supply structure that is standard to ACDR and various other firm-level trade models, the demand side influences the gains from trade through two channels: i) the shape of the function D determines markups and the markup elasticity ρ ; ii) the demand shifters Q and F affects price levels through the choke price.

Using the expenditure function, we obtain the changes in welfare for country j :

$$d \log e_j = (1 - \rho) \sum_i \lambda_{ij} d \log(w_i \tau_{ij}) + \rho d \log P_j \quad (12)$$

The first term is identical to ACDR and captures the direct effect of the changes in marginal costs and markups, while the second term captures the indirect effect on price levels through the choke price, which differs from ACDR as shown above. With $\sum_i \lambda_{ij} d \log(w_i \tau_{ij}) = \frac{1}{\theta} d \log \lambda_{jj}$ (as in Arkolakis et al., 2012), the gains from trade $dGT_j = -d \log e_j$ can be summarized by the following proposition:

Proposition 2 *For small changes in trade costs, the changes in welfare for country j are given by:*

$$dGT_j = - \left(\frac{1 - \rho}{\theta} + \frac{\rho \varepsilon_{Fj}}{(\theta + 1) \varepsilon_{Fj} - \varepsilon_{Qj}} \right) d \log \lambda_{jj} \quad (13)$$

Conditional on changes in import penetration $d \log \lambda_{jj}$, trade elasticity θ and markup elasticity ρ , the range of admissible elasticities ε_{Fj} and ε_{Qj} is such that the gains from trade can take any value between $-\frac{1-\rho}{\theta} d \log \lambda_{jj}$ (indirectly-additive case) and $-\frac{1}{\theta} d \log \lambda_{jj}$ (homothetic case).

This includes the following three cases:

$$\begin{aligned} dGT_j &= -\frac{1}{\theta} d \log \lambda_{jj} && \text{with homothetic preferences (ACR, ACDR)} \\ dGT_j &= -\left(\frac{1-\rho}{\theta} + \frac{\rho}{\theta+1}\right) d \log \lambda_{jj} && \text{Directly-separable preferences (ACDR)} \\ dGT_j &= -\frac{1-\rho}{\theta} d \log \lambda_{jj} && \text{Indirectly-separable preferences (Berlotti et al, 2018)} \end{aligned}$$

By shutting down the effect on the choke price, Bertolotti et al (2018) obtain the lowest bound.

Again, it is practical to assume $F(\Lambda) = \Lambda$ and $Q(\Lambda) = \alpha \Lambda^\beta$. Gains from trade then satisfy the same formula as in ACDR:

$$dGT_j = - \left(\frac{1 - \rho}{\theta} + \frac{\rho}{\theta + 1 - \beta} \right) d \log \lambda_{jj} \quad (14)$$

but now with a continuous representation in β , where β can take any value between $-\infty$ and 1 (and potentially higher if it remains smaller than the price elasticity ε_D), implying again that lower and upper bounds correspond to the indirectly-additive and homothetic cases.

4 Empirical implementation

Without any prior on the shape of preferences, this generalization leaves us with at least an additional parameter to estimate if we want to quantify the gains from trade, focusing on the the iso-elastic case described in equation (14). Estimates of other parameters, the trade elasticity θ and the markup elasticity ρ , can be borrowed from previous estimation approaches given that most of them are consistent with the current framework. For instance, Simonovska and Waugh (2014) provide an estimate of the trade elasticity θ ; Amiti et al. (2014) and Arkolakis et al (2018), among many others, discuss how to identify the markup elasticity ρ , which determines the pass-through of costs to prices.

The new parameter β influences the gains from trade through the choke price in particular, and determines how price levels depends on income and trade costs. Hence, to estimate β , a natural approach is to examine how prices set by exporting firms depend on marginal costs as well as characteristics of the destination country such as income and trade costs.

As described previously, the price of a product is determined by its marginal cost and its markup, which can itself be expressed as a function of the marginal cost and the choke price. A firm with a marginal cost draw c would charge a price $p_{ijt}(c) = \tau_{ijt}w_{it}c\mu(P_{jt}/(\tau_{ijt}w_{it}c))$. Differentiating, i.e. considering small changes in costs and choke prices over time, we obtain:

$$\Delta \log p_{ijt}(c) = (1 - \rho_{ij}(c))\Delta \log(\tau_{ijt}w_{it}) + \rho_{ij}(c) \Delta \log P_{jt} \quad (15)$$

The choke price P_j is not directly observed, but its relation to trade costs can be indirectly inferred through the gravity equation. Differentiating equation (7), we get:

$$\Delta \log P_{jt} = \frac{1 - \beta}{1 + \theta - \beta} \Delta \log w_{jt} + \frac{1}{1 + \theta - \beta} \Delta \log MA_{jt} \quad (16)$$

where MA_{jt} refers to “market access” for destination j : $MA_{jt} \equiv \sum_i N_i b_i w_i^{-\theta} \tau_{ij}^{-\theta}$. This measure of market access has been often used in the literature and can be structurally estimated using gravity equations, taking the sum of estimated exporter fixed effects adjusted by trade costs.⁹ Next, taking averages across firms in country i and combining equations (15) and (16) leads to:

$$\Delta \log p_{ijt} = (1 - \rho)\Delta \log(\tau_{ijt}w_{it}) + \frac{\rho}{1 + \theta - \beta} \Delta \log MA_{jt} + \frac{\rho(1 - \beta)}{1 + \theta - \beta} \Delta \log w_{jt} \quad (17)$$

This relationship that can be used to identify both ρ and β , assuming that income is not perfectly correlated with market access and that the econometrician observes exogenous variations in market access (e.g. as discussed in Feyrer, 2009).

In this regression, the first coefficient describes the pass-through of costs to prices, which can be used to identify the pass-through elasticity ρ , as often done in practice. In turn, the second term can be used to identify $\frac{\rho}{1 + \theta - \beta}$. A smaller β is associated with a lower sensitivity of prices to market access in the destination country, and a higher sensitivity to income levels. For instance, in the limit case

⁹See e.g. Redding and Venables (2004), Head and Mayer (2010), and Fally et al. (2010) among others.

where preferences are indirectly-additive and $\beta = -\infty$, the coefficient for market access is null. In the general case, this coefficient can be used to estimate the second term in equation (14), $\frac{\rho}{\theta+1-\beta} \Delta \log \lambda_{jkt}$, and thus estimate the gains from trade with generalized separability.

5 Conclusion

This paper examines the implications of trade for consumer welfare by considering a flexible yet tractable demand system (Gorman, 1972, 1995, Pollak, 1972, Fally, 2018) encompassing directly-separable, indirectly-separable and homothetic preferences with a single aggregator. Embedded in a general-equilibrium trade model featuring gravity, such demand system allows for more flexible effects of trade and income on prices. An upper bound of the gains from trade is provided by the homothetic case (as in Arkolakis et al., 2012), while a lower bound corresponds to the case of indirectly-separable preferences (as in Bertolotti et al., 2018), allowing for a continuum of intermediate cases.

In particular, the added flexibility is reflected in how trade affects choke prices (reservation prices). To estimate the gains from trade and identify additional parameters that characterize the demand system, a suggestion is to exploit how changes in trade costs affect markups using firm-level data on exports from a specific source country.

References

- Amiti, M., O. Itskhoki, and J. Konings (2014). Importers, exporters, and exchange rate disconnect. *American Economic Review* 104(7), 1942–78.
- Arkolakis, C., A. Costinot, D. Donaldson, and A. Rodríguez-Clare (2015). The elusive pro-competitive effects of trade. NBER Working Paper 21370.
- Arkolakis, C., A. Costinot, and A. Rodríguez-Clare (2012). New trade models, same old gains? *American Economic Review* 102(1), 94–130.
- Behrens, K., G. Mion, Y. Murata, and J. Südekum (2014). Trade, wages, and productivity. *International Economic Review* 55(4), 1305–1348.
- Bertolotti, P. and F. Etro (2017). Monopolistic competition, as you like it. *University Ca' Foscari of Venice, Dept. of Economics Research Paper Series, 08/2017*.
- Bertolotti, P., F. Etro, and I. Simonovska (2018). International trade with indirect additivity. *American Economic Journal: Microeconomics* 10(2), 1–57.
- Chaney, T. (2008). Distorted gravity: the intensive and extensive margins of international trade. *American Economic Review* 98(4), 1707–21.
- Costinot, A. and A. Rodríguez-Clare (2018). The US gains from trade: Valuation using the demand for foreign factor services. *Journal of Economic Perspectives* 32(2), 3–24.
- De Loecker, J. and F. Warzynski (2012). Markups and firm-level export status. *American economic review* 102(6), 2437–71.
- Fally, T. (2018). Integrability and generalized separability. NBER Working Paper 25025.

- Fally, T., R. Paillacar, and C. Terra (2010). Economic geography and wages in brazil: Evidence from micro-data. *Journal of Development Economics* 91(1), 155–168.
- Feenstra, R. C. (2018a). Alternative sources of the gains from international trade: Variety, creative destruction, and markups. *Journal of Economic Perspectives* 32(2), 25–46.
- Feenstra, R. C. (2018b). Restoring the product variety and pro-competitive gains from trade with heterogeneous firms and bounded productivity. *Journal of International Economics* 110, 16–27.
- Feyrer, J. (2009). Trade and income-exploiting time series in geography. Technical report. NBER Working Paper 14910.
- Gorman, W. M. (1972). Conditions for generalized additive separability. Unpublished transcript, now printed in Gorman (1995).
- Gorman, W. M. (1995). *Collected Works of WM Gorman: Separability and Aggregation*, Volume 1. Oxford University Press.
- Head, K. and T. Mayer (2010). Gravity, market potential and economic development. *Journal of Economic Geography* 11(2), 281–294.
- Hummels, D. and P. J. Klenow (2005). The variety and quality of a nation’s exports. *American Economic Review* 95(3), 704–723.
- Matsuyama, K. and P. Ushchev (2017). “Beyond CES: Three alternative classes of flexible homothetic demand systems”. *Working paper*.
- Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71(6), 1695–1725.
- Pollak, R. A. (1972). Generalized separability. *Econometrica: Journal of the Econometric Society*, 431–453.
- Redding, S. and A. J. Venables (2004). Economic geography and international inequality. *Journal of international Economics* 62(1), 53–82.
- Simonovska, I. (2015). Income differences and prices of tradables: Insights from an online retailer. *The Review of Economic Studies* 82(4), 1612–1656.
- Simonovska, I. and M. E. Waugh (2014). The elasticity of trade: Estimates and evidence. *Journal of international Economics* 92(1), 34–50.

Appendix

A) Effect of income on the price aggregator

Differentiating equation (1) yields:

$$\int_{\omega \in \Omega} \lambda(\omega) \left[\varepsilon_Q \frac{d \log \Lambda}{d \log w} - \varepsilon_D \varepsilon_{Fj} \frac{d \log \Lambda}{d \log w} + \varepsilon_D \right] d\omega = 1$$

We obtain equation (3) by solving for $\frac{d \log \Lambda}{d \log w}$. Note that ε_Q and ε_F depend only on Λ while ε_D varies across product varieties and depend on $F(\Lambda)p(\omega)/w = p(\omega)/P$.

B) Gravity equation

Marginal costs are given by $w_i c$ where c is drawn from a cumulative distribution $G_i(c) = b_i c^\theta$. Firms with draws above $c_{ij}^* = P_j / (w_i \tau_{ij})$ do not sell to market j . As in ACDR, it is then convenient to make a change in variable and define $v = c_{ij}^* / c$ with a cumulative distribution $1 - v^{-\theta}$ with the adjusted mass of firms $N_i b_i P_j^\theta (w_i \tau_{ij})^{-\theta}$. For firms in country i exporting to j with marginal cost $c = c_{ij}^* / v = P_j / (w_i \tau_{ij} v)$, their price is $P_j \mu(v) / v$ and their sales are given by:

$$\tilde{x}(v) Q(\Lambda_j) P_j = \mu(v) / v D(\mu(v) / v) Q(\Lambda_j) P_j$$

where $\mu(v)$ is defined implicitly by equation (5). Summing up across all firms in country i , total exports to country j correspond to:

$$X_{ij} = N_i b_i P_j^\theta (w_i \tau_{ij})^{-\theta} \int_{v=1}^{+\infty} \tilde{x}(v) Q(\Lambda_j) P_j \theta v^{-\theta-1} dv = \chi N_i b_i (w_i \tau_{ij})^{-\theta} Q(\Lambda_j) P_j^{1+\theta}$$

with with constant term $\chi \equiv \int_{v=1}^{+\infty} \mu(v) / v D(\mu(v) / v) \theta v^{-\theta-1} dc$. This gives equation (6) in the text.

C) Proposition 1

Differentiating equation (7) w.r.t Λ_j yields:

$$\begin{aligned} (\varepsilon_{Qj} - (\theta + 1)\varepsilon_{Fj}) d \log \Lambda_j - \theta \sum_i \lambda_{ij} d \log(w_i \tau_{ij}) &= 0 \\ \Rightarrow d \log \Lambda_j &= -\frac{\theta}{(\theta + 1)\varepsilon_{Fj} - \varepsilon_{Qj}} \sum_i \lambda_{ij} d \log c_{ij} \end{aligned}$$

and thus we obtain equation (8) in the text for $d \log P_j = -\varepsilon_{Fj} d \log \Lambda_j$. Next, we can apply the same trick as in Arkolakis et al. (2012) to obtain:

$$\begin{aligned} \sum_i \lambda_{ij} d \log(w_i \tau_{ij}) &= \frac{1}{\theta} \sum_i \lambda_{ij} d \log(w_i \tau_{ij})^\theta = \frac{1}{\theta} \sum_i \lambda_{ij} d \log(\lambda_{jj} / \lambda_{ij}) \\ &= \frac{1}{\theta} d \log \lambda_{jj} - \frac{1}{\theta} \sum_i \lambda_{ij} d \log \lambda_{ij} = \frac{1}{\theta} d \log \lambda_{jj} - \frac{1}{\theta} \sum_i d(\lambda_{ij}) = \frac{1}{\theta} d \log \lambda_{jj} \end{aligned}$$

D) Proposition 2

Given that we normalize income $w_j = 1$ in country j , welfare changes are given by the change in the expenditure function. Applying Shephard's Lemma, this is equal to the change in log prices weighted by expenditure shares:

$$\begin{aligned} d \log e_j &= \sum_i \int_{\omega \in \Omega_i} \lambda_{ij}(\omega) d \log p_{ij}(\omega) \\ &= (1 - \rho) \sum_i \lambda_{ij} d \log(w_i \tau_{ij}) + \rho d \log P_j \end{aligned}$$

where ρ is the sales-weighted average of the markup elasticity defined in equation (11). We obtain Proposition 2 as a corollary of Proposition 1 combined with the welfare change above.