Modeling Demand in International and Macro Economics

by

Daniel Patrick Murphy

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Public Policy and Economics)
in the University of Michigan
2013

Doctoral Committee:
  Professor Lutz Kilian, Co-Chair
  Professor Alan Deardorff, Co-Chair
  Professor John Bound
  Assistant Professor Andrei Levchenko

DEDICATION

To Aiko, Maura, my parents (Donald and Susan), and my siblings (Shannon, Chris, and Michelle).

ACKLOWLEDGEMENTS

I would first like to thank my co-chairs, Alan Deardorff and Lutz Kilian, whose time, dedication, and encouragement were invaluable to my research. I had the great pleasure of working with Lutz on a couple of projects, and am incredibly grateful for the mentorship that grew out of our time spent working together. In addition to being a wonderful teacher and mentor, Lutz was a willing sounding board for my ideas while encouraging me to pose focused research questions that fit into the context of an established literature. He put in countless hours helping me to develop these ideas and to improve the exposition of my papers.

A number of my research ideas over the past few years were non-traditional or outside the boundaries of my field of international economics. Alan encouraged me to pursue and develop these ideas, and it is impossible to understate the value of his encouragement. I am profoundly grateful for his time, his mentorship, and his insights. I am a far better economist because of Alan, and I feel incredibly fortunate to have Alan as a co-chair.

I would also like to thank Andrei Levchenko for his input and advice. Andrei's suggestions greatly improved my first chapter, and I am especially grateful for his advice on how to write my paper to an audience of international economists, as well as how to navigate the publication process. John Bound also provided helpful advice on how to fit some of my ideas into the context of an established literature.

My wife, Aiko, was a wonderful support. At times I struggled with how to merge my ideas with contemporary economic thought, and her encouragement and understanding were crucial, especially during times when the sense of discouragement felt overwhelming. My parents likewise have always encouraged my desire to pursue higher education, and their support has been tremendous.

Over the past few years I've met a number of times with Alasdair MacIntyre at Notre Dame. Alasdair approaches economics from the perspective of a philosopher/sociologist and encouraged me to investigate economics-related questions that are often overlooked by academic economists. My conversations with Alasdair helped inspire my research agenda on the nature and implications of consumer demand.

Finally, I would like to acknowledge the influence of Denver East High School. My high school years were intellectually formative; I learned how to think and to question, and much of my passion to improve as a social scientist is a result of my experiences at East and the intellectual environment fostered by the people there.

TABLE OF CONTENTS

CHAPTER

LIST OF TABLES

TABLE

LIST OF FIGURES

FIGURE

CHAPTER I

Why are Goods and Services more Expensive in Rich Countries? Demand Complementarities and Cross-Country Price Differences

Abstract

Empirical studies show that tradable consumption goods are more expensive in rich countries. This paper proposes a simple yet novel explanation for this apparent failure of the law of one price: Consumers' utility from tradable goods depends on their consumption of complementary goods and services. Monopolistically competitive firms charge higher prices in countries with more complementary goods and services because consumer demand is less elastic there. The paper embeds this explanation within a static Krugman (1980)-style model of international trade featuring differentiated tradable goods. Extended versions of the model can account for the high prices of services in rich countries, as well as for several stylized facts regarding investment rates and relative prices of investment and consumption across countries. The paper provides direct evidence in support of this new explanation. Using free-alongside-ship prices of U.S. and Chinese exports, I demonstrate that prices of specific subsets of tradable goods are higher in countries with high consumption of relevant complementary goods, conditional on per capita income and other country-level determinants of consumer goods prices.

1. Introduction

There is abundant evidence that tradable goods are more expensive in countries with high per-capita incomes. In particular, recent studies of disaggregate data on tradable goods show a failure of the law of one price due to firms charging higher markups for goods sold to rich countries than for goods sold to poor countries. For example, Alessandria and Kaboski (2011) find that rich countries pay more for goods leaving U.S. docks, and Simonovska (2011) documents that an online apparel retailer charges higher markups to consumers in rich countries.[1]

This paper proposes a simple explanation to account for this evidence: The utility a consumer derives from tradable goods depends on his consumption of other goods and services that complement the tradable goods. Higher utility from tradable goods lowers the price elasticity of demand for tradables, causing monopolistically competitive firms to charge higher markups in markets with high consumption of complementary goods and services. Since consumers in rich countries can afford more complementary goods and services, they have a lower price-elasticity of demand for tradable consumer goods and are charged higher prices for tradables.

One example of such a complementary good is housing, which complements the demand for consumer tradables such as a home entertainment system. In the U.S., consumers have relatively inelastic demand for home entertainment systems because they also have spacious TV rooms in their homes and a reliable supply of energy. In Ecuador, in contrast, the average consumer has less space in his home and an unreliable power supply. Firms can therefore charge a higher price in the U.S. than in Ecuador for identical entertainment systems.

Demand for new consumer goods also depends on public infrastructure, including roads and public safety. The value of a car, for example, depends not only on features specific to the vehicle, but also on the environment in which the car is driven. Paved roads increase the utility from owning a nice car, as does a safe environment with low probability of the car being stolen, while owning the same car may provide far less utility in an area with dirt roads or in an area that is insecure.

Many types of goods and services may complement demand for differentiated consumer goods (and differentiated consumer goods could complement demand for each other). To distinguish the complementary goods from the consumer goods in the analysis below, I refer to these complementary goods and services as *catalyst* goods. Often catalyst goods will be durables, such as housing or public infrastructure, but they may also be services or intangibles,

---

[1] Additional empirical work corroborates this evidence of a failure of the law of one price for tradables. Gopinath, Gourinchas, Hsieh, and Li (2011) demonstrate that wholesalers charge different markups in the U.S. market than in the Canadian market. Fitzgerald and Haller (2012) and Burstein and Jaimovich (2008) also find that wholesale prices differ substantially across destinations, even when the products are made in the same plant. Their evidence suggests that cross-country price differences are driven by characteristics specific to the destination countries.

such as public safety, or other consumer goods. The concept of a catalyst captures the notion that some goods and services facilitate consumers' derivation of utility from other final goods and services. The notion of catalysts is similar to the notion of consumer demand proposed by Lancaster (1966), who suggests that goods and services are not direct objects of utility themselves but rather contain properties and characteristics that consumers combine to generate utility.

This explanation based on *demand complementarity and pricing-to-market* is simple, but to my knowledge has not been explored to date.[2] Below I embed this explanation within a general equilibrium model that builds on a class of utility functions developed in the trade literature that yield demand curves with nonconstant price elasticities of demand. The baseline model features demand complementarity between catalyst goods and differentiated final consumption goods. Specifically, the intercept of the demand curve for a differentiated final good depends on the level of consumption of catalyst goods. Section 2 develops the basic intuition within a closed economy and demonstrates that as the country's income increases, it consumers more catalyst goods and pays higher prices for differentiated consumer goods. Section 3 extends the analysis to two countries with the aim of explaining the relevant empirical facts with respect to prices of tradable goods across countries. In equilibrium, the rich country consumes more catalyst goods and pays more for tradable goods.

Sections 4 and 5 extend the model to demonstrate that the mechanism responsible for the high tradable prices in rich countries can also account for a number of other stylized facts in the trade and growth literatures. Section 4 incorporates nontraded services into a two-country model and shows that complementarities between catalyst goods and nontraded services also generate high prices of nontraded services.[3] The typical explanation for the observed correlation between country per-capita income and nontradable prices is based on the theory developed by Harrod (1933), Balassa (1964), and Samuelson (1964), collectively referred to as HBS. The HBS model postulates that the law of one price holds in tradables, and that rich-country productivity is higher in the tradable sector than in the nontradable sector. High productivity in the tradable sector

---

[2] The term *pricing-to-market* refers to general price discrimination across countries. Krugman (1987) defines pricing-to-market as price discrimination in response to nominal exchange rate movements. A number of authors since then, including Alessandria and Kaboski (2011), refer to the term more generally.

[3] The positive relationship between prices of nontradables and income is well documented. See, for example, Balassa (1964), Samuelson (1964), Baghwati (1984), Summers and Heston (1991), Barro (1991), Hsieh and Klenow (2007), and Alessandria and Kaboski (2011).

drives up wages in rich countries, which causes higher prices in the sector with lower productivity (nontradables).

As recently noted by Alessandria and Kaboski (2011), it is unlikely that HBS can fully explain the price-income relationship across countries because the difference between tradable-sector productivity and nontradable-sector productivity within rich countries is too small to account for the strong relationship between prices and incomes across countries. In contrast to HBS, the explanation proposed in Section 4 for the high price of nontradables in rich countries does not rely on sectoral productivity differentials. Rather, the driving mechanism is complementarity between catalyst goods (e.g., housing, roads, public safety, or any other complementary good) and final goods and services, which causes monopolistically competitive firms in the tradable and nontradable sectors to charge higher markups when a country has more catalyst goods.

The model extensions in Sections 2 through 4 are static and thus abstract from differences in the durability of different goods, and from the accumulation of capital for production. Nonetheless, some of the goods that are considered catalysts (e.g. housing and roads) are, in reality, more durable than final goods (e.g. electronics). Furthermore, while housing and roads are fixed assets that are not traded, they are produced using traded investment goods. Sections 2 through 4 abstract from these complications for the sake of simplicity and because doing so has no bearing on the basic mechanism driving the model. Section 5 demonstrates that incorporating these additional dimensions of reality can help explain why real investment rates are low in poor countries.

Hsieh and Klenow (2007) show that (1) investment goods are no more expensive at international prices in poor countries, and (2) real investment as a fraction of GDP per capita is positively correlated with income per capita. Based on these observations, and on the fact that consumption is more expensive in countries with high per capita income, Hsieh and Klenow conclude that poor countries must have lower productivity in their tradable consumption goods sector than in their nontraded goods sector. This conclusion leads them to declare a "productivity puzzle": Why are poor countries even worse at producing tradable consumption goods than they are at producing consumption services? Hsieh and Klenow challenge the literature to explain this apparent productivity differential in poor countries.[4]

The extended model in Section 5 matches the empirical regularities highlighted by Hsieh and Klenow without relying on sectoral productivity differences in poor countries. The mechanism driving the results, demand complementarity and pricing-to-market, is the same mechanism responsible for the high price of consumption goods in rich countries in Sections 3 and 4. Furthermore, in the same way that demand complementarity and pricing-to-market provides an alternative to the HBS-based conclusion that rich countries must have a sectoral productivity differential, it also provides an alternative to Hsieh and Klenow's hypothesis of a poor country productivity differential.

An important question is whether the explanation proposed in this paper fits the micro data. Section 6 of the paper provides independent empirical evidence that prices of consumer goods depend on a country's consumption of the relevant catalyst goods and services. Specifically, I use U.S. and Chinese export data to investigate whether certain consumer goods are sold at higher prices to countries with higher stocks of relevant catalysts. I show that household goods and electronic goods are sold at higher prices to countries with more housing and electricity, conditional on per capita income and other country-level determinants of consumer goods prices. Also, new cars are sold at higher prices to countries with higher percentages of paved roads.

Simonovska (2011) is the most closely related paper that offers an explanation for high prices of tradables in rich countries.[5] In Simonovska's model, high tradable prices in rich countries are due to low demand elasticities (and corresponding high markups) arising from consumption of larger varieties of imported goods. In the models above, high prices reflect high consumption of catalyst goods, rather than differences in the set of imported goods. Furthermore, demand complementarity and pricing-to-market causes high prices in a closed economy setting as well as in an open economy setting and can account for a number of empirical regularities in the trade and growth literatures. Thus, while both the model in

---

[4] Buera, Kaboski, and Shin (2011) suggest that one reason for the productivity differential in poor countries is that manufacturing requires economies of scale, which must be supported by a well-developed financial sector. Poor countries face financial frictions which disproportionately lower manufacturing productivity (and hence productivity in the investment good sector).

[5] Hummels and Lugovsky (2009) and Alessandria and Kaboski (2011) also propose theoretical explanations for the positive correlation between markups and income per capita.

Simonovska (2011) and the model here employ forms of nonhomothetic preferences that permit price-dependent demand elasticities, the underlying mechanisms are different. One implication of the demand-complementarities explanation is that the extent to which markups vary across countries should depend on the extent to which the tradable good in question is complementary to other goods and services.

The demand-side explanation for high prices of tradable goods in rich countries explored here complements a burgeoning literature that examines demand-side explanations for the cross-country relationship between income and quality of imports.[6] Fajgelbaum, Grossman, and Helpman (2011) develop a model featuring complementarity between a homogenous good and quality of vertically differentiated goods. In their model, higher incomes are associated with more purchases of higher quality goods, but not with higher markups paid for those goods. An interesting avenue for future research is to develop models in which high consumption of catalyst goods is associated with purchases of higher quality goods *and* higher markups for a good of any given quality.

2. Closed Economy Model

This section illustrates in a closed-economy setting how prices of consumer goods increase with a country's wealth due to markups that rise with the country's stock of catalyst goods. The closed economy features a representative consumer with preferences over differentiated final goods, a homogenous catalyst good, and a homogenous numeraire good. The final goods represent appliances, household items, and cars, among other consumer goods. The homogenous catalyst good represents housing and public infrastructure such as roads, energy supply, safety, and any other good that may complement demand for the final goods.

The catalyst is produced under perfect competition by a representative firm, while the consumer goods are produced by monopolistically competitive firms. Both sectors use labor, which is supplied inelastically, as the only factor of production. The numeraire is endowed to the economy and enters the consumer's utility function linearly. This particular setup is based on a variant of the linear demand system developed by Ottaviano, Tabuchi, and Thisse (2002), and

---

[6] This paper more broadly fits into work that explores the implications of nonhomothetic preferences for patterns of trade, including Bergstrand (1990), Hunter (1991), Matsuyama (2000), Mitra and Trindade (2005) and Fieler (2011), among many others. Markusen (2010) reviews the literature and discusses a range of phenomena for which non-homothetic preferences improve the correspondence between trade models and the data.

is chosen to demonstrate in the simplest possible setting how demand complementarities and pricing-to-market cause prices of final consumer goods to rise with a country's wealth. The Ottaviano et al (2002) demand system is analytically convenient, in part because the marginal utility of income is unity for all levels of income. Appendix I.A demonstrates that the results of this section are robust to alternative specifications for which the marginal utility of income varies with income and the numeraire is produced with labor.

Model Setup. The representative agent's utility function is defined over the catalyst good C, the mass Ω of final goods, and a numeraire y:

$$U = y + C^{\alpha} \int_{\Omega} f_{\omega} d\omega - \frac{1}{2}\gamma \int_{\Omega} f_{\omega}^{2} d\omega, \qquad (1)$$

where f_{ω} is consumption of final good $\omega \in \Omega$. The numeraire y is endowed to the economy, and could represent any commodity, such as gold or wheat. Agricultural commodities are perhaps the most intuitive interpretation of the numeraire because, among other reasons, agriculture is often considered to be endowed to the economy due to its heavy reliance on immobile factors of production.[7]

Equation (1) is a simplified version of the utility functions used in Ottaviano et al (2002), Melitz and Ottaviano (2008), and Foster, Haltiwanger, and Syverson (2008). The utility function here differs from their utility functions in two ways. First, the marginal utility from consuming any variety ω is independent of consumption of any other variety $\omega' \neq \omega$. This is for analytical convenience only. Second, equation (1) features a catalyst good C that acts as a demand shifter for the consumption goods.

The agent inelastically supplies L units of labor to the market. The agent also owns the firms in the economy and receives profit income from the mass Ω of firms that produce differentiated consumption goods. The budget constraint is

$$y + wL + \int_{\Omega} \Pi_{\omega} d\omega = y + p_{C}C + \int_{\Omega} p_{\omega} f_{\omega} d\omega, \qquad (2)$$

where w is the wage, p_C is the price of the catalyst, and p_{ω} is the price of variety ω.

Maximizing (1) subject to (2) yields demand for final good f_{ω}:

$$f_{\omega}^{d} = \frac{1}{\gamma}(C^{\alpha} - p_{\omega}), \qquad (3)$$

---

[7] See, for example Ottaviano et al (2002), and, more recently, Allen (2012) for models with an endowed agricultural commodity.

which is increasing in C. This simple linear demand function captures the notion that demand for consumption goods is less elastic when the economy has a higher stock of housing and public infrastructure. For example, a consumer's willingness to pay for a fancy new oven is higher (and his price-sensitivity lower) if he has a nice kitchen and house that can accommodate dinner guests.

Demand for the catalyst is likewise increasing in consumption of final goods:

$$C^d = \left(\frac{\alpha F}{p_C}\right)^{\frac{1}{1-\alpha}}, \tag{4}$$

where $F \equiv \int_0^{\Omega} f_\omega d\omega$. The larger the mass of goods Ω, and the more of each good consumed, the higher is the demand for the catalyst. For example, demand for a mansion is high if a consumer has access to artwork, furniture, and appliances with which to fill the mansion. Otherwise a large, empty house is of little value.

Final Good Sector. Final good firms employ labor in a linear production function to produce output according to

$$f_\omega = AL_\omega, \tag{5}$$

where A is labor productivity, which is identical across firms and across sectors, and L_ω is the amount of labor employed by firm ω. Each firm chooses its output price to maximize profits. Firm ω's profit function is

$$\Pi_\omega = p_\omega f_\omega - \frac{w}{A} f_\omega. \tag{6}$$

The profit-maximizing price is derived by substituting (3) into (6) and maximizing with respect to p_ω:

$$p_\omega = \frac{1}{2}\left(C^\alpha + \frac{w}{A}\right). \tag{7}$$

Prices are increasing in C because demand is less elastic when C is high. Equation (7) captures the intuition that (a) monopolistically competitive firms charge a price that is proportional to consumer utility from consumption of firms' output, and (b) catalyst goods increase utility from consumption of final goods. The two-country counterpart to (7) in Section 3 derives the central result that rich countries pay higher prices for identical goods. Note that linearity of the demand curve (3) is sufficient but not necessary for the price elasticity of demand to be decreasing in the catalyst. Appendix I.B derives the necessary and sufficient conditions under which the price elasticity of demand is decreasing in C.

Given the price, demand for variety ω is

$$f_{\omega}^{d} = \frac{1}{2\gamma}\left(C^{\alpha} - \frac{w}{A}\right), \tag{8}$$

which is derived by substituting (7) into (3). Firm ω earns profits given by

$$\Pi_{\omega} = \frac{1}{4\gamma}\left(C^{\alpha} - \frac{w}{A}\right)^{2}.$$

I permit profits to be positive because incorporating a zero-profit condition would simply complicate the model by adding an equilibrium equation and an extra endogenous variable (the mass of final goods firms). Also, abstracting from fixed costs and increasing returns permits a clear comparison of productivity across sectors to demonstrate that demand complementarities, rather than productivity differentials, drive the price differences in the two-country models in sections 3 through 5. Nonetheless, the positive relationship between final goods prices and economic wealth derived below is robust to incorporating zero profits as a long-run equilibrium condition.

Since productivity is identical across firms, so are prices and quantities: $f_{\omega} = f$ and $p_{\omega} = p \ \forall \ \omega \in \Omega$. Total demand over all final consumption goods is derived by integrating (8) across varieties:

$$F = \frac{\Omega}{2\gamma}\left(C^{\alpha} - \frac{w}{A}\right). \tag{9}$$

Given total demand for final goods, we can write demand for labor in the final good sector as $L_Q \equiv \int_0^{\Omega} L_{\omega} d\omega$, or

$$L_Q = \frac{1}{A}F. \tag{10}$$

Catalyst Sector. Catalysts are produced competitively using the technology

$$C = AL_C, \tag{11}$$

where L_C is labor in the catalyst sector. Cost minimization yields the price of catalysts, $p_C = w/A$.

Equilibrium. Equilibrium is characterized by demand for catalysts (4), demand for consumer goods (9), and labor market clearing,

$$L = \frac{1}{A}(F + C). \tag{12}$$

The endogenous variables are $F, C,$ and w.

Comparative Statics. The central message of this section is that in general equilibrium, markups and prices of final goods are increasing in the economy's wealth. Figure I.1 shows how market outcomes vary with productivity under the following parameterization:

$$L = 1, \qquad A = 1, \qquad \Omega = 1, \qquad \alpha = 0.3, \qquad \gamma = 0.3.^{8} \tag{13}$$

As A increases, the price of the catalyst falls and the quantity of the catalyst increases. The increase in C shifts out the demand curve for final goods, lowering the price-elasticity of demand. Firms charge a higher markup, causing a higher price of final goods. The positive effect of C on demand for final goods outweighs the counteracting effect of the increase in w on the price, so overall demand for final goods increases. Thus, even in this simple closed economy, prices and quantities of final goods rise with economy-wide productivity due to high demand from the consumption of more catalyst goods.

3. Two-Country Model

This section extends the model of Section 2 to incorporate trade between two countries N (North) and S (South). The purpose of this exercise is to demonstrate that demand complementarities and pricing-to-market can account for the evidence of higher prices of tradable goods in rich countries than in poor countries. In the model, each country is endowed with the numeraire and inelastically supplies labor to produce catalyst goods and differentiated final goods. Catalyst goods are not traded across countries. This assumption is for simplicity (the qualitative results are robust to permitting the catalyst to be traded), and because some catalyst goods represent housing and public infrastructure, which are fixed immobile assets. The numeraire is endowed to each country and is traded. Following Krugman (1980), each country specializes in a unique set of differentiated final goods. As in Section 2, final goods are produced by monopolistically competitive firms. Firms can move final goods costlessly across international borders. Consumers, however, face large costs of moving goods across international borders. Therefore even though firms charge country-specific prices, consumers do

---

[8] The qualitative results with respect to the markup are robust to all parameter values. A proof based on total differentiation of the equilibrium equations is available from the author upon request.

not arbitrage because there are prohibitive costs associated with doing so. These costs could represent the time required to travel across international borders, as suggested in Gopinath et al (2011), as well as other transportation costs and information rigidities.

Model Setup. Each country $j \in \{N, S\}$ produces a mass Ω_j of final goods which are consumed at home and abroad. Goods produced in country j are indexed by $\omega_j \in \Omega_j$. The utility function of the representative consumer in country j is

$$U_j = y_j + \sum_{i=N,S} \int_{\omega_i \in \Omega_i} \left[C_j^{\alpha} f_j(\omega_i) - \frac{\gamma}{2}\left(f_j(\omega_i)\right)^2 \right] d\omega_i, \tag{14}$$

where y_j and C_j are consumption of the numeraire and catalyst by country j and $f_j(\omega_i)$ is consumption in country j of variety ω_i from country $i \in \{N, S\}$. As in the previous section, the numeraire good y simplifies the analysis.

The budget constraint of the representative agent in country j is

$$y_j^0 + w_j L_j + \sum_{i=N,S} \int_{\omega_j \in \Omega_j} \Pi_i(\omega_j) = y_j + p_{Cj} C_j + \sum_{i=N,S} \int_{\omega_i \in \Omega_i} p_j(\omega_i) f_j(\omega_i), \tag{15}$$

where y_j^0 is the endowment of the numeraire in country j, $\Pi_i(\omega_j)$ is the profit from sales of variety ω_j to country i, y_j is the amount of the numeraire consumed in country j, p_{Cj} is the price of the catalyst in j, and $p_j(\omega_i)$ is the price of variety ω_i in j.

Consumer optimization with respect to $f_j(\omega_i)$ yields demand for variety ω_i in country j:

$$f_j^d(\omega_i) = \frac{1}{\gamma}\left(C_j^{\alpha} - p_j(\omega_i)\right). \tag{16}$$

Similarly, the first order condition with respect to C_j yields

$$C_j^d = \left(\frac{\alpha F_j}{p_{Cj}}\right)^{\frac{1}{1-\alpha}}, \tag{17}$$

where $F_j \equiv \sum_{i=N,S} \int_{\omega_i \in \Omega_i} f_j(\omega_i)$ is the total quantity of final goods consumed in country j.

Consumption Good Sector. Output in the final goods sector is produced using the technology

$$f(\omega_j) = A_j L_{\omega_j}, \tag{18}$$

where $f(\omega_j) \equiv f_N(\omega_j) + f_S(\omega_j)$. Each firm ω_j charges a country-specific price to maximize the profits $\Pi_i(\omega_j)$ from selling variety ω_j in country $i \in \{N, S\}$. I assume that if

$p_S(\omega_S) \neq p_N(\omega_S)$, the costs to consumers in country $\{i: p_i(\omega_S) < p_j(\omega_S)\}$ of purchasing good ω_S in i are sufficiently high to prevent arbitrage. Likewise, costs to consumers of transporting good ω_N across international borders are sufficiently high to prevent arbitrage when $p_S(\omega_N) \neq p_N(\omega_N)$.

Profits from sales of ω_j in i can be written

$$\Pi_i(\omega_j) = p_i(\omega_j) f_i(\omega_j) - \frac{w_j}{A_j} f_i(\omega_j). \tag{19}$$

The profit-maximizing price charged in country i is

$$p_i(\omega_j) = \frac{1}{2}\left(C_i^\alpha + \frac{w_j}{A_j}\right). \tag{20}$$

Equation (20) states that the optimal price of an identical good varies across countries based on the stock of catalyst goods in each country. This is the key result of the paper, and it explains why rich countries pay higher prices for tradable goods. Of course, it remains to be seen that rich countries have more of the catalyst in equilibrium, a task to which we now turn.

Given the price defined by (20), consumer demand in country i for ω_j is

$$f_i^d(\omega_j) = \frac{1}{2\gamma}\left(C_i^\alpha - \frac{w_j}{A_j}\right), \tag{21}$$

The resulting revenues of firm ω_j from sales to country i are

$$p_i(\omega_j) f_i(\omega_j) = \frac{1}{4\gamma}\left(C_i^{2\alpha} - \frac{w_j^2}{A_j^2}\right), \tag{22}$$

and profits are

$$\Pi_i(\omega_j) = \frac{1}{4\gamma}\left(C_i^\alpha - \frac{w_j}{A_j}\right)^2. \tag{23}$$

Catalyst Sector. As in Section 2, the catalyst in country j is produced competitively according to $C_j = A_j L_{Cj}$, where A_j is productivity in country j and L_{Cj} is labor employed in j's catalyst sector. The price of the catalyst is $p_{Cj} = w_j/A_j$, which is derived from cost minimization by the representative catalyst firm. Since the catalyst is not traded across countries, there is no role for comparative advantage and each country will produce some of the catalyst in equilibrium.

Equilibrium. Since $p_i(\omega_j)$ and $f_i(\omega_j)$ are identical for any variety ω_j from country j, it will be helpful to omit variety indices by writing $p_{ij} = p_i(\omega_j)$, $f_{ij} = f_i(\omega_j)$, and $\Pi_{ij} =$

$\Pi_i(\omega_j)\ \ \forall\, \omega_j \in \Omega_j$. Then F_j becomes $F_j = \Omega_j f_{jj} + \Omega_i f_{ji}$. The budget constraint in country j simplifies to

$$y_j^0 + w_j L_j + \Omega_j(\Pi_{jj} + \Pi_{ij}) = y_j + p_{Cj} C_j + \Omega_j p_{jj} f_{jj} + \Omega_i p_{ji} f_{ji} \qquad (24)$$

Labor market clearing in j is $L_j = L_{Qj} + L_{Cj}$, where $L_{Qj} \equiv \int_{\omega_j \in \Omega_j} L_{\omega_j} d\omega_j$ is total labor used in the final goods sector. By substituting in the production functions for final goods and the catalyst, labor market clearing in country j can be written

$$L_j = \frac{1}{A_j}\left(\Omega_j(f_{jj} + f_{ij}) + C_j\right). \qquad (25)$$

Market clearing for the numeraire is

$$y_N^0 + y_S^0 = y_N + y_S. \qquad (26)$$

Equilibrium is characterized by demand for the catalyst in each country (17), demand for final goods (21), labor market clearing in each country (25), market clearing for the numeraire (26), and the budget constraints (24). By Walras' Law, one of these equations is redundant. For clarity, the equilibrium conditions are written explicitly as:

$$C_N = \left(\frac{A_N \alpha (\Omega_N f_{NN} + \Omega_S f_{NS})}{w_N}\right)^{\frac{1}{1-\alpha}}, \qquad C_S = \left(\frac{A_S \alpha (\Omega_N f_{SN} + \Omega_S f_{SS})}{w_S}\right)^{\frac{1}{1-\alpha}},$$

$$f_{NN} = \frac{1}{2\gamma}\left(C_N^\alpha - \frac{w_N}{A_N}\right), \quad f_{NS} = \frac{1}{2\gamma}\left(C_N^\alpha - \frac{w_S}{A_S}\right),$$

$$f_{SS} = \frac{1}{2\gamma}\left(C_S^\alpha - \frac{w_S}{A_S}\right), \quad f_{SN} = \frac{1}{2\gamma}\left(C_S^\alpha - \frac{w_N}{A_N}\right),$$

$$L_N = \frac{1}{A_N}[\Omega_N(f_{NN} + f_{SN}) + C_N], \qquad L_S = \frac{1}{A_S}[\Omega_S(f_{SS} + f_{NS}) + C_S],$$

$$y_N^0 + y_S^0 = y_N + y_S,$$

$$y_N^0 - y_N + \Omega_N p_{SN} f_{SN} = \Omega_S p_{NS} f_{NS},$$

where the last equilibrium equation is a simplified version of the budget constraint for country N (see equation 24). The ten equations above yield a unique solution for the endogenous variables $w_N, w_S, y_N, y_S, C_N, C_S, f_{NN}, f_{NS}, f_{SS},$ and f_{SN}.

Results. Figure I.2 shows relative prices in N and S of identical goods under the following baseline parameterization:

$$A_N, A_S = 3, \qquad L_N, L_S, y_N^0, y_S^0 = 1, \qquad \Omega_S, \Omega_N = 0.5, \qquad \alpha, \gamma = 0.3$$[9]

The left-hand graph shows the ratio of prices relative to the numeraire, while the graph on the right shows the ratio of PPP-adjusted prices.[10] According to Figure I.2, the model predicts that as a country gets richer, it pays higher prices for identical goods than does its poorer counterpart, consistent with the evidence across countries cited in the introduction. Specifically, goods produced in N are more expensive in N, and goods produced in S are more expensive in N.

Figure I.3 shows how market outcomes vary with productivity in N. As A_N rises, N produces and consumes more of the catalyst. Higher catalyst consumption shifts out the demand curves of final goods, which causes firms from both countries to charge higher markups for goods sold in N. The resulting quantities of final goods demanded by N increase because the outward shift of the demand curves caused by higher catalyst consumption outweighs the movement along the demand curves caused by higher prices. Therefore a rise in A_N causes higher catalyst and final good consumption in N, as well as higher prices of final goods.

The rise in f_{NS} requires S to devote more labor resources to its export sector and less resources to production for domestic consumption, causing a fall in f_{SS} and C_S. How is this optimal for S? Since exports from S are sold at a higher markup, the value of exports f_{NS} increase relative to the value of the numeraire. S therefore reallocates labor to the export sector to exchange for the numeraire and for consumer goods produced in N, leading to an increase in trade and an increase in welfare in S. Figure I.4 shows that welfare in both countries increases with A_N.

Summary of the Two-Country Model. As productivity in N increases, N can afford to produce more catalyst goods, which shifts out its demand for final goods by increasing the price-

---

[9] Baseline productivity is set to 3 to ensure that utility from consumption of final goods and catalyst goods is sufficiently high to ensure positive demand for imports from N and S. In other words, the productivity parameters are chosen such that the equilibrium is at an interior solution given by the ten equilibrium equations above.

[10] In Figure I.4, $p_{ij}^R \equiv p_{ij}/P_i$, where P_i is the consumer price index. P_i is normalized to unity under the initial calibration in which productivity is equal across countries. Note that PPP holds when N and S are equal because the exact same bundles are purchased at identical costs in each country. When productivity is not equal across countries (e.g. at any point in Figure I.4 to the right of the y-axis), P_i is the current price in country i of the bundle of goods consumed when PPP held (the Laspeyres Index):

$$P_i = \frac{y_i^0 + p_{Ci}C_i^0 + p_{ij}\Omega_j f_{ij}^0 + p_{ii}\Omega_i f_{ii}^0}{y_i^0 + p_{Ci}^0 C_i^0 + p_{ij}^0\Omega_j f_{ij}^0 + p_{ii}^0\Omega_i f_{ii}^0},$$

where the superscript 0 indicates the price or quantity that prevails when PPP holds (productivity is equal across countries).

intercept of the demand curve. N's resulting lower price elasticity of demand causes firms to charge a higher markup in N than in S, which increases relative prices in N.

As we will see in Section 4 below, this simple explanation of demand complementarity and pricing-to-market can explain not only high prices of traded consumer goods in rich countries, but also high prices of nontradables in rich countries.

4. Two-Country Model with Nontradables

This section extends the model of Section 3 to incorporate nontradables that are produced and sold domestically by monopolistically competitive firms. The purpose of this simple extension is to demonstrate that the mechanism emphasized above to account for the comparatively high prices of tradables in rich countries can also account for the comparatively high prices of nontradables in rich countries.[11]

The typical explanation for the observed correlation between country income per capita and nontradable prices is based on the theory developed by Harrod (1933), Balassa (1964), and Samuelson (1964). The HBS model assumes that the law of one price (LOP) holds in tradables, and that rich-country productivity is higher in the tradable sector than in the nontradable sector. High productivity in the tradable sector drives up wages in rich countries, which causes higher prices in the sector with lower productivity (nontradables).

As recently noted by Alessandria and Kaboski (2011), there are at least two strong reasons to doubt HBS as a full explanation of the price-income correlation across countries. First, the LOP does not hold for tradables, violating a key assumption of HBS. Second, the rise in relative productivity of tradables within rich countries appears too small to account for the strong relationship between prices and incomes across countries.

The model extension below provides an alternative explanation to account for comparatively high prices of nontradables in Rich countries (as well as comparatively high tradable prices). In contrast to HBS, the new explanation does not rely on sectoral productivity differentials. Rather, the driving mechanism is complementarity between catalyst goods and final goods, as in Section 3. Rich countries can afford to produce more catalyst goods, which in turn increases demand for nontradable goods and services.

---

[11] This high price of nontradables in rich countries is well-documented. See, for example, Alessandria and Kaboski (2011, p.92).

Consider, for example, purchasing car rental services in Ecuador, which has unpaved roads and a generally unsafe environment for driving. Even if a car rental agency can provide a vehicle to rent at low cost, customers will have low preference for this service simply because there are characteristics specific to Ecuador (poor driving conditions) which may not affect the cost to the firm of providing the service, but which reduce customers' utility from the service. Likewise, consumers may require a haircut once a month, but the utility from a haircut at a barber shop relative to cutting one's own hair depends on the convenience of traveling to the barber, which in turn depends on public infrastructure such as roads, safety, and reliable energy supply to ensure the barber shop will be open for business. It may also depend on the prevalence of other goods and services for which one might need a haircut to fully enjoy. Salon services are more valuable, for example, when consumers attend formal events in which a certain style of appearance is the cultural norm. Notice that in this last example, the complementary *catalyst* is itself a service.

Utility from nontradable services also depends on durables, such as housing. For example, the value of services such as window-washing, carpet-cleaning, and lawn mowing all depend on whether consumers have homes that can accommodate windows, carpets, and lawns. In Quito, Ecuador, these services are of little value because few homes there are suitable for windows and nice carpets, and few households own lawns.

The model below captures this intuition by incorporating nontradable services into the model from Section 3. As we will see, high service prices will rely on demand complementarities, rather than on sectoral productivity differentials.

Model Setup. The representative consumer in country j has utility over the numeraire, tradable goods, and a mass Ψ_j of nontradable services:

$$\begin{aligned} U_j = y_j &+ \int_{\psi_j \in \Psi_j} \left[C_j^{\alpha} f_j(\psi_j) - \frac{\gamma}{2}\left(f_j(\psi_j)\right)^2 \right] d\psi_j \\ &+ \sum_{i=N,S} \int_{\omega_i \in \Omega_i} \left[C_j^{\alpha} f_j(\omega_i) - \frac{\gamma}{2}\left(f_j(\omega_i)\right)^2 \right] d\omega_i, \end{aligned} \tag{27}$$

where ψ_j indexes the services in country j and $f_j(\psi_j)$ is consumption of variety ψ_j in j. Unless otherwise stated, the notation and variable names are the same as in Section 3 above.

Country j's budget constraint is

$$\begin{aligned} y_j^0 + w_j L_j + \int_{\psi_j \in \Psi_j} \Pi_\mathrm{j}(\psi_j) + \sum_{i=N,S} \int_{\omega_j \in \Omega_j} \Pi_i(\omega_j) \\ = y_j + p_{Cj} C_j + \int_{\psi_j \in \Psi_j} p_j(\psi_j) f_j(\psi_j) + \sum_{i=N,S} \int_{\omega_i \in \Omega_i} p_j(\omega_i) f_j(\omega_i), \end{aligned} \tag{28}$$

where $\Pi_\mathrm{j}(\psi_j)$ are profits from sales of service ψ_j at price $p_j(\psi_j)$ and quantity $f_j(\psi_j)$. Consumer optimization with respect to $f_j(\psi_j)$ yields demand for variety ψ_j in country j:

$$f_j^d(\psi_j) = \frac{1}{2\gamma}\left(C_j^\alpha - p_j(\psi_j)\right). \tag{29}$$

Demand for tradable goods is given by (16) above, and demand for catalyst goods is given by (17), where total consumption of final goods and services in country j is

$$F_j \equiv \int_{\psi_j \in \Psi_j} f_j(\psi_j) + \sum_{i=N,S} \int_{\omega_i \in \Omega_i} f_j(\omega_i).$$

Final Goods Firms. Optimization by firms in the tradable sector is identical to that in section 3. As above, prices and quantities are independent of the variety, so we can write

$$p_{ji} = \frac{1}{2}\left(C_j^\alpha + \frac{w_i}{A_i}\right), \tag{30}$$

$$f_{ji} = \frac{1}{2\gamma}\left(C_j^\alpha - \frac{w_i}{A_i}\right), \tag{31}$$

where p_{ji} and f_{ji} are the price and quantity of any variety produced in country $i \in \{N,S\}$ and sold in $j \in \{N,S\}$.

Service Sector Firms. Services are produced using the same technology as that used by consumer goods:

$$f_j(\psi_j) = A_j L_{\psi_j}, \tag{32}$$

where L_{ψ_j} is labor used to produce service variety ψ_j. Profits of firm ψ_j are

$$\Pi_\mathrm{j}(\psi_j) = p_j(\psi_j) f_j(\psi_j) - \frac{w_j}{A_j} f_j(\psi_j). \tag{33}$$

Profit maximization yields the price

$$p_j(\psi_j) = \frac{1}{2}\left(C_j^\alpha + \frac{w_j}{A_j}\right). \tag{34}$$

The resulting quantity demanded is

$$f_j(\psi_j) = \frac{1}{2\gamma}\left(C_j^\alpha - \frac{w_j}{A_j}\right). \tag{35}$$

Catalyst Sector. The production function for catalyst goods is $C_j = A_j L_{Cj}$. As in Sections 2 and 3, the catalyst sector is perfectly competitive. The price of the catalyst is $p_{Cj} = w_j / A_j$, which is derived from cost minimization by the representative catalyst firm. Also, as in Sections 2 and 3, the catalyst is not traded across countries.

Equilibrium. Since $p_j(\psi_j)$ and $f_j(\psi_j)$ are identical for any variety ψ_j from country j, it is helpful to omit variety indices by writing $p_j = p_j(\psi_j)$ and $f_j = f_j(\psi_j)\ \forall\, \psi_j \in \Psi_j$. Total consumption of goods and services in country j can be written $F_j = \Psi_j f_j + \Omega_j f_{jj} + \Omega_i f_{ji}$.

Equilibrium is characterized by demand for the catalyst in each country (17), demand for final goods in each country (31), demand for nontradables in each country (29), labor market clearing in each country,

$$L_N = \frac{1}{A_N}[\Psi_N f_N + \Omega_N (f_{NN} + f_{SN}) + C_N],$$

$$L_S = \frac{1}{A_S}[\Psi_S f_S + \Omega_S (f_{NS} + f_{SS}) + C_S],$$

numeraire market clearing

$$y_N^0 + y_S^0 = y_N + y_S,$$

and the budget constraint for N, which simplifies to

$$y_N^0 + \Omega_N p_{SN} f_{SN} = y_N + \Omega_S p_{NS} f_{NS}.$$

By Walras' Law, the budget constraint in S is redundant.

Results. The initial parameter values are $\alpha = 0.3$, $\gamma = 0.3$; the initial productivity paramaters are set to unity, and the mass of goods and services in each country is unity ($\Psi_j + \Omega_j + \Omega_i = 1$). Figure I.5 shows market outcomes as productivity in N increases. The results are very similar to those from Section 3: N's production and consumption of catalyst and final goods increases, as does the price of tradables in N. In addition, the relative price of services is higher in N because the increase in C_N lowers the price elasticity of demand for services, causing service-sector firms in N to charge a higher markup than service-sector firms in S.

Summary of Two-Country Model with Services. The value of services within a country rises with that country's stock of catalyst goods. A rich country can afford to produce more of the catalyst, which lowers the price elasticity of demand for tradable final goods and nontradable services within the country. As a result, monopolistically competitive firms in the final good and service sectors charge a higher markup, causing higher prices of tradable goods and nontradable services in the rich country.

Note that the simple mechanism of demand complementarity and pricing-to-market was initially proposed in Sections 2 and 3 to account for the high prices of tradable goods in rich countries. Section 4 showed how the same mechanisms can account for another stylized fact in international trade (the high prices of services in rich countries) by adding a degree of realism to the baseline model. Of course, even the extended model of Section 4 abstracts from many dimensions of reality. One of the most obvious abstractions is the absence of traded investment goods. As we will see in Section 5, incorporating traded investment goods and capital as a factor of production permits the model to explain additional stylized facts in the trade and growth literatures.

Section 5: Incorporating Capital and Tradable Investment.

Hsieh and Klenow (2007) highlight the following empirical regularities in the growth literature:

1) The price of consumption is high when income per capita is high.
2) Prices of investment goods are no higher in poor countries.
3) Real investment rates are positively correlated with income.

The first fact, which is typically attributed to HBS, has already been discussed at length. Fact (3) dates back to Barro (1991) and is often attributed to policies in poor countries that distort savings and investment decisions. Hsieh and Klenow provide evidence in support of fact (2) and search for a unified explanation of the facts. They conclude,

> "Poor countries appear to have low investment rates in PPP terms primarily because they have either low productivity in producing investment goods or low productivity in producing tradables to exchange for investment goods…Our results thus imply…a deeper productivity puzzle. The challenge is to explain not only low overall productivity in poor countries, but also low productivity in investment goods (or in providing consumption goods to trade for investment goods) relative to consumption goods" (p. 564, emphasis mine).

The reader is referred to Hsieh and Klenow (2007) for why they infer a productivity puzzle based on the three empirical regularities.[12] This section provides an alternative unified explanation for these stylized facts that does not rely on poor countries having low productivity in the investment goods sector relative to productivity in the consumption goods sector. My explanation instead builds on the mechanisms developed above under a framework that features pricing-to-market in the final goods sector and complementarity between catalyst goods and final goods.

More concretely, in the model below goods are produced using labor, which is inelastically supplied, and capital, which is accumulated through investment. Rich countries have a high stock of catalyst goods, which causes the value of final consumer goods to be higher in rich countries than in poor countries and causes higher markups for goods sold to Rich countries (fact 1).

Fact (2) is an immediate consequence of any assumption on the market structure for investment goods such that prices of investment goods equalize across countries. In the simplest case, investment goods are produced under perfect competition (as in Hsieh and Klenow 2007) and are traded costlessly. An alternative assumption is that differentiated investment goods are produced by monopolistically competitive firms. If the differentiated goods are aggregated into the investment good through a CES aggregator, then firms will charge the same markup over marginal cost in each country for their investment good and the price of the final investment good will equalize across countries.

Hsieh and Klenow note that under some empirical specifications, investment goods are slightly more expensive in rich countries. A model in which the investment aggregator function gives rise to price-dependent investment demand curves can generate a positive relationship between investment prices and income, as demonstrated in Appendix I.D. The properties of such a model are more complicated than is necessary to demonstrate that the focal mechanism, demand complementarity and pricing-to-market, can resolve Hsieh and Klenow's productivity puzzle. The model in this section presents the simplest case of perfect competition in the investment goods sector, consistent with the analysis in Hsieh and Klenow (2007). Appendix I.C

---

[12] Recently, Valentinyi and Herrendorf (2012) estimate that developing countries' TFP in tradable manufactured goods is about equal to average TFP, which suggests that an explanation other than productivity differentials is required to explain facts (1) through (3).

demonstrates the case in which a final investment good is produced from differentiated intermediate investment goods using a CES aggregator.

In the two-country model below, the homogenous investment good is traded costlessly, causing the price of the investment good to equalize across countries. This implies that the rental rate of capital also equalizes across countries, consistent with the evidence in Caselli and Feyrer (2007) that marginal products of capital are similar across countries. The equalization of capital prices across countries causes the capital price-to-wage ratio to be high in poor countries relative to the ratio in rich countries. In response to the difference in factor prices, firms in poor countries demand a lower capital/labor ratio than do rich countries, which lowers real investment in poor countries relative to investment in rich countries (fact 3).

All goods are produced using a Cobb-Douglas technology that employs labor and capital as factor inputs. The homogenous investment good is traded, as are differentiated final goods. The catalyst (e.g. housing and infrastructure) is not traded. Since the catalyst represents durables such as housing and roads, as well as nondurables that may complement consumer goods, the catalyst is permitted to be long-lived in the model.

The price of investment is equalized across both countries, so the country with a comparative advantage in the investment sector will produce the investment good, while the other country will trade final consumer goods for the investment good. The homogenous capital investment good is not produced in S because economy-wide productivity in S is assumed to be low enough that S is better off exchanging consumption goods for investment goods.[13] This assumption approximates reality: Eaton and Kortum (2001) show that poor countries import most of their capital equipment. Finally, the model abstracts from production of nontradable final goods and services for the sake of simplicity only.

Model Setup. The representative consumer in country $j \in \{N, S\}$ maximizes

$$\sum_{t=0}^{\infty} \beta^t U_{jt}$$

subject to

---

[13] To rule out the possibility of a within-country productivity differential, I assume that the South has access to the technology to produce investment goods using the same total factor productivity as in other sectors and verify that in equilibrium they are better off producing consumption goods to exchange for investment goods.

$$K_{j,t+1} = (1-\delta)K_{jt} + I_{jt},$$

$$C_{j,t+1} = (1-\delta)C_{jt} + X_{jt},$$

$$y_{jt}^{0} + R_{jt}K_{jt} + w_{jt}L_{jt} + \sum_{i=N,S}\int_{\omega_j\in\Omega_j}\Pi_{it}(\omega_j)$$
$$= p_{Ijt}I_{jt} + y_{jt} + p_{Xjt}X_{jt} + \sum_{i=N,S}\int_{\omega_i\in\Omega_i} p_{jt}(\omega_i)f_{jt}(\omega_i),$$

where U_{jt} is the within-period utility function given by (14), K_{jt} is the capital stock in period $t\in\{0,1,2,\ldots\}$, R_{jt} is the rental price of capital, I_{jt} is capital investment by j in period t, p_{Ijt} is the price of capital investment, X_{jt} is the addition to j's catalyst stock in period t, p_{Xjt} is the price of X_{jt}, and δ is depreciation of capital and the catalyst. The remaining variables are as defined in Section 3.

The analysis carried out here is in steady state, so from now on time subscripts will be omitted. Consumer optimization with respect to K yields the steady-state rental price of capital:

$$R = p_I(r+\delta), \qquad (36)$$

where $r = \frac{1-\beta}{\beta}$ is the real interest rate. Since investment is traded at no cost, its price equalizes across countries ($p_{Ij} = p_I$), as does the rental price of capital.

Steady-state demand for the catalyst in country j is

$$C_j = \left[\frac{\beta\alpha F_j}{p_{Xj}\big(1-\beta(1-\delta)\big)}\right]^{\frac{1}{1-\alpha}}, \qquad (37)$$

where

$$F_j \equiv \sum_{i=N,S}\int_{\omega_i\in\Omega_i} f_j(\omega_i).$$

Demand for consumer good variety ω_i in country j is given by (16).

Consumption Good Sector. Output in the consumption goods sector is produced using the technology

$$f(\omega_j) = A_j L_{\omega j}^{\eta} K_{\omega j}^{1-\eta}, \qquad (38)$$

where $f(\omega_j) \equiv f_N(\omega_j) + f_S(\omega_j)$ an A_j is total factor productivity in each sector in country j. As in the baseline model, each firm ω_j charges a country-specific price to maximize the profits $\Pi_i(\omega_j)$ from selling variety ω_j in country $i\in\{N,S\}$.

Also, costs to consumers of transporting goods across international borders are sufficiently high to prevent arbitrage.

The profit-maximizing price charged in country i is

$$p_i(\omega_j) = \frac{1}{2}\left(C_i^\alpha + \frac{c_j}{A_j}\right), \tag{39}$$

where

$$c_j = \frac{1}{\eta^\eta (1-\eta)^{1-\eta}}\, w_j^\eta R^{1-\eta}$$

is the cost-minimizing price of a unit of output at unit total factor productivity.
Equation (39) is the Section 5 counterpart to equation (20), and it accounts for the high price of consumer goods in rich countries.

Given the price defined by (39), consumer demand in country i for ω_j is

$$f_i^d(\omega_j) = \frac{1}{2\gamma}\left(C_i^\alpha - \frac{c_j}{A_j}\right). \tag{40}$$

Catalyst Investment Sector. Catalyst investment in country j is produced under perfect competition according to

$$X_j = A_j L_{Xj}^\eta K_{Xj}^{1-\eta}. \tag{41}$$

The price of catalyst investment is $p_{Xj} = c_j/A_j$. Since the catalyst investment good is not traded across countries, there is no role for comparative advantage and each country will produce some catalyst investment in equilibrium.

Capital Investment Sector. Capital investment is produced in country N under perfect competition according to

$$I = A_N L_I^\eta K_I^{1-\eta}. \tag{42}$$

The price of capital investment is $p_I = c_N/A_N$. Country N purchases some of the investment good and exports the rest. Market clearing implies

$$I = I_N + I_S. \tag{43}$$

Equilibrium. I solve for fifteen unknowns,

$$w_N, w_S, y_N, y_S, C_N, C_S, f_{NN}, f_{NS}, f_{SS}, f_{SN}, p_{NS}, p_{SN}, R, K_N, K_S,$$

using the following fifteen equilibrium conditions:

$$C_N = \left[\frac{\alpha A_N(\Omega_N f_{NN} + \Omega_S f_{NS})}{c_N(r+\delta)}\right]^{\frac{1}{1-\alpha}} \qquad C_S = \left[\frac{\alpha A_S(\Omega_S f_{SS} + \Omega_N f_{SN})}{c_S(r+\delta)}\right]^{\frac{1}{1-\alpha}}$$

$$L_N = \left(\frac{R}{w_N}\frac{\eta}{1-\eta}\right)^{1-\eta}\left[\frac{\Omega_N(f_{NN}+f_{SN})}{A_N}+\frac{X_N}{A_N}+\frac{I_N+I_S}{A_N}\right].$$

$$L_S = \left(\frac{R}{w_S}\frac{\eta}{1-\eta}\right)^{1-\eta}\left[\frac{\Omega_S(f_{SS}+f_{NS})}{A_S}+\frac{X_S}{A_S}\right]$$

$$y_N^0 - y_N + \frac{c_N}{A_N}I_S + \Omega_N p_{SN} f_{SN} = \Omega_S p_{NS} f_{NS}$$

$$y_N^0 + y_S^0 = y_N + y_S$$

$$R = w_N\left(\frac{1}{A_N}\frac{r+\delta}{\eta^\eta(1-\eta)^{1-\eta}}\right)^{\frac{1}{\eta}}$$

$$f_{NN} = \frac{1}{2\gamma}\left(C_N^\alpha - \frac{c_N}{A_N}\right) \quad f_{SN} = \frac{1}{2\gamma}\left(C_S^\alpha - \frac{c_N}{A_N}\right) \quad f_{SS} = \frac{1}{2\gamma}\left(C_S^\alpha - \frac{c_S}{A_S}\right) \quad f_{NS} = \frac{1}{2\gamma}\left(C_N^\alpha - \frac{c_S}{A_S}\right)$$

$$p_{NS} = \frac{1}{2}\left(C_N^\alpha + \frac{c_S}{A_S}\right) \qquad p_{SN} = \frac{1}{2}\left(C_S^\alpha + \frac{c_N}{A_N}\right),$$

$$K_N = \frac{w_N}{R}\frac{1-\eta}{\eta}L_N \qquad K_S = \frac{w_S}{R}\frac{1-\eta}{\eta}L_S \tag{44}$$

where

$$c_N = \frac{w_N}{\eta}L_N \qquad c_S = \frac{w_S}{\eta}L_S$$

$$I_N = \delta K_N \qquad I_S = \delta K_S.$$

Results. Figure I.6 shows relative prices and investment under the following initial parameter values:

$$A_N = 4, \qquad A_S = 2, \qquad y_N^0, y_S^0 = 3, \qquad L_N, L_S, \Omega_N, \Omega_S = 1, \qquad \alpha, \gamma = 0.3,$$

$$\beta = 0.99, \qquad \delta = 0.3.$$

Recall that productivity in S is identical across sectors, which excludes the possibility discussed in Hsieh and Klenow (2007) of productivity differentials in S driving the results. Even though S does not produce the investment good in equilibrium, it is assumed to have access to the technology to produce the investment good using the same total factor productivity as prevails in the other sectors.

As in the Section 3, the rich country, N, pays more for final goods due to a lower price elasticity of demand stemming from higher consumption of the catalyst. N also purchases more of the investment good because its ratio of the capital price to the wage is lower than the corresponding ratio in S. This is because high productivity in N causes w_N to be high relative to w_S. Demand for capital in each country is given by (44). Since labor supply is equal across countries, and $\frac{w_N}{R} > \frac{w_S}{R}$, demand for capital (and investment goods) is higher in N. As shown in Figure I.6, actual investment and the real investment rate are higher in N than in S.

Summary of Model with Investment. Hsieh and Klenow (2007) infer from facts (1) through (3) above that poor countries must be worse at producing investment goods (which are primarily tradable) than at producing consumption goods (which include a substantial nontraded component). Their hypothesis of a productivity differential in poor countries is a corollary of the Harrod-Balassa-Samuelson hypothesis in the sense that, under their proposed explanation, poor countries have lower productivity in a primarily tradable sector (investment) than in a primarily nontraded sector (consumption).

This section proposes an alternative explanation for the facts based on demand complementarities and pricing-to-market: High levels of catalysts in the rich country cause a high real wage and high consumption prices there. Since investment prices equalize across countries (due either to perfect competition, constant markups in a monopolistically competitive investment sector, or complete cross-country capital markets), the rental rate on capital also equalizes across countries. The high wage-to-rent ratio in the rich country causes high demand for capital goods there.

Implications. A shortcoming that is shared by the Harrod-Balassa-Samuelson hypothesis and the Hsieh-Klenow hypothesis is that it is not intuitively clear why productivity should differ across sectors within a country to the extent required to explain the observed price patterns. The mechanism I propose, demand complementarities and pricing-to-market, is based on intuitive consumption patterns and the realistic assumption that firms have market power. That a single intuitive mechanism can provide a unified explanation for a number of puzzles in the trade and growth literatures is attractive from a modeling point of view, but we also need to ask how compatible this mechanism is with the micro data. The next section provides independent empirical evidence in support of this mechanism's relevance for observed price patterns.

6. Empirical Evidence

So far I have emphasized the ability of a single mechanism, demand complementarity and pricing-to-market, to account for a number of cross-country stylized facts. Here I test the dependence of consumer prices on countries' consumption of catalyst goods using data on U.S. and Chinese exports. The challenge in the empirical work is to distinguish the effect of demand complementarities from other mechanisms that may cause a positive correlation between consumer prices and income per capita across countries. Indeed, income and catalyst consumption are perfectly correlated in the theoretical models above, and if the same were true of reality it would be impossible to distinguish between demand complementarities and other potential explanations for the price-income relationship. In reality, however, catalyst consumption is imperfectly correlated with income per capita, which permits me to test the dependence of prices on the component of catalyst consumption that is not correlated with income.[14]

The analysis in this section examines three catalyst goods in particular: electricity, housing, and roads. Each of these catalyst goods is an imperfect correlate with GDP per capita, and each is expected to be a strong complement for a different subset of consumer goods. Electricity complements demand for electric goods, houses complement demand for household goods (e.g. televisions and furniture), and roads complement demand for new cars. Therefore, the model predicts the following, conditional on country-level fixed effects:

1) Electric goods are sold at higher prices in countries with a more reliable power supply (or superior energy infrastructure).
2) Household goods are sold at higher prices in countries with more housing per capita.
3) New cars are more expensive in countries with better roads.

To explore these predictions, I obtain prices of goods sold to different countries from disaggregated data on U.S. and Chinese exports. The U.S. Exports Harmonized System data, available on Robert Feenstra's webpage, contains unit values and quantities of bilateral exports leaving US docks for each Harmonized System (HS)-10 product category. As discussed by Alessandria and Kaboski (2011), there are two advantages of using this data to study the extent of pricing-to-market for tradable goods. First, the disaggregated nature of the data mitigates

---

[14] There are many potential reasons for the imperfect correlation between catalyst consumption and income. I do not suggest any particular reason, but assume that these reasons are exogenous to prices of consumer imports.

potential concerns that different unit values may reflect differences in quality. Second, the unit values are free-alongside-ship values, which exclude transportation costs, tariffs, and additional costs incurred in the importing country.

To test the three hypotheses it is necessary to identify 'household goods', 'electric goods', and 'new cars' separately from other consumer goods. This task is fairly straightforward for new cars, which I classify as any good for which its HS-10 description indicates that it is a new passenger vehicle. Identification of electric goods is also fairly straightforward, although some goods are not identified as electric but require electricity to use (such as a television). I classify as 'electric' any consumer good (end-use code 40000-50000) that is labeled as electric and not battery-powered, as well as a number of clearly electric goods, including TVs, stereos, and associated parts.

Classifying household goods is more difficult because most consumer goods are stored in homes. Nonetheless, some goods are more directly complementary to housing than others. Consider a house with an extra bedroom and bathroom. The extra space is likely to complement demand for furniture, bedding, towels, and similar goods. Also, a country with more homes per capita will have more need for kitchen items. Therefore I classify all furniture, glassware, chinaware, cookware, cutlery, tools, rugs, TVs, VCRs, and stereo equipment (end-use codes 41000, 41010, 41020, 41040, 41200, and 41210) as household goods. I also classify appliances (end-use 14030) as household goods, with the exception of air conditioners and radiators, the demand for which I assume depends more on weather than on housing. Other goods such as clothing and personal care items are excluded from the list of household goods because they are not directly complementary to housing. Table I.1 lists the subset of consumer goods that I classify as household goods.

To corroborate the evidence from U.S. export data, I test hypotheses 1 and 2 using Chinese Customs export data, which contain free-alongside-ship values and quantities of goods at the HS-8 level of disaggregation.[15] Despite the lower level of disaggregation, the Chinese dataset has a number of advantages over the U.S. export data. First, the dataset contains identifiers for firms and firm locations, which help control for quality variation within a product category. Second, China exports far more consumer goods to a broader range of countries. The

---

[15] I am incredibly grateful to Jagadeesh Sivadasan and Michael Olabisi for sharing the Chinese Export data. I do not test the third hypothesis using the Chinese data because the dataset does not include and exports of new passenger vehicles in 2005.

Chinese dataset does not have end use codes or descriptions, so I identify consumer, household, and electronic good HS-8 categories as those categories that contain only consumer goods, only household goods, and only electronic goods as HS-10 subcategories.

Country-level data on the catalyst goods are from the International Comparison Program (ICP) and the World Development Indicators at the World Bank. Heston (2011) provides the ICP's measures of the dwelling services for Europe in 2005. The measure of the dwelling services in Europe is based on a survey of rental rates, from which the ICP assigned countries an index of their per capita housing volume. Measures of housing volume in other regions are either unreliable, or are not comparable to the measure of housing in Europe (see Heston 2011 for a discussion).

I use electricity consumption as a proxy for a country's energy infrastructure. Country-level data on electricity consumption per capita are from the World Development Indicators at the World Bank. The measure of a country's road quality is the percent of roads that are paved, also available from the World Development Indicators. Most countries do not have data on road quality for more than a single year between 2002 through 2006, so I pick the most recent year for which data is available as a country's measure of road quality.

I test the three hypotheses outlined earlier separately in the following subsections.

6.1 Electricity Infrastructure and Prices of Electric Goods

First, I assess whether prices of exports of electric goods depend on countries' access to electricity. As a proxy for a country's electricity access, I use data on electricity consumption per capita, provided by the World Development Indicators. This proxy is most appropriate in underdeveloped countries with low average electricity consumption per capita. In developed countries, differences in electricity consumption are more likely to reflect differences in factors other than the population's access to electricity, such as weather. Therefore I limit my attention to countries that consumed less than 5 mega-watt-hours of electricity per person in 2005. This restriction removes most European countries from the sample, as well as other wealthy countries such as Japan and Qatar, and leaves 72 countries in the sample. Portugal, South Africa, and Malta are the remaining countries with the highest per capita electricity consumption.

I test the following empirical specification:

$$p_{ch} = \alpha_c + \gamma_h + \psi q_{ch} + \beta \text{MWHpercap}_c \text{Egood}_h + \epsilon_{ch}, \tag{45}$$

where p_{ch} is the log of the unit value of good h exported to country c, normalized by its within-good standard deviation.[16] The coefficient α_c represents country fixed effects, γ_h represents fixed effects for each good category, and q_{ch} is the log quantity of good h sold to country c, normalized by its within-good standard deviation. MWHpercap_c is the per capita electricity consumption in country c, Egood_h indicates whether good h is electric, and ϵ_{ch} denotes the regression error. Unit values and quantities for each country-product pair in the U.S. data are averages of the values between 2004 and 2006 (the three most recent years available).[17] The Chinese data are only available in 2005. To prevent nonrepresentative products from driving the results, the samples are limited to country-product pairs with over 100 units sold and to products that are exported to at least 10 countries.

The coefficient β captures the extent to which the markup for electric goods depends on electricity access. β can be interpreted as representing a causal relationship if electricity consumption is exogenous to the product price. Electricity consumption is indeed likely to be exogenous with respect to the price of a single imported product. If there is any endogenous response to electric prices, equations (16) and (17) imply electricity consumption should respond negatively to high import prices. In this case, high electricity consumption is associated with low prices of electric goods, and β will underestimate the causal effect of access to electricity on electric goods prices. In other words, the estimate of β is biased downward in the presence of endogenous electricity consumption.[18]

I include quantity as a regressor in (45) to capture the dependence of firms' costs on the quantity they sell to a given destination. A negative estimate for ψ may reflect bulk discounts, or other cost savings from repeated transactions between U.S. sellers and foreign buyers. Omitting

---

[16] When the regression is run on Chinese data, the price is normalized by its standard deviation within a firm-product pair. This normalization prevents goods with large price dispersion from driving the results, and mitigates potential concerns that the regression results may be driven by differences in quality. Manova and Zhang (2012), for example, document that Chinese firms that charge a wide range of prices for their exports also pay a wide range of prices for imported inputs. They infer on the basis of this evidence that these firms sell goods of varying quality.

[17] Averaging unit values across time has the advantage of averaging out the noise in the yearly data while preserving the ability to identify β based on the cross-sectional variation across destination countries. When the regression is run on yearly data (rather than averaged data), the results are similar but with slightly larger standard errors.

[18] As a robustness check, I used 2002 values of electricity consumption as an instrument and obtained nearly identical results to those presented below. This is unsurprising given that electricity consumption in 2005 is nearly perfectly correlated with electricity consumption in prior years.

quantity would bias downward β to the extent that higher electricity-related demand for electric goods is associated with higher quantities sold and lower marginal costs.

More generally, conditioning on quantity controls for demand parameters and cost parameters that may vary across country-product pairs. Monopolistically competitive firms charge a price that depends on catalyst consumption as well as other demand and cost parameters. Since these parameters may vary across countries in a way that is correlated with catalyst consumption, conditioning on quantity controls for these parameters and permits an interpretation of β as the partial effect on the price of an increase in catalyst consumption, conditional on a country's position on its demand curve.[19]

Table I.2 shows the estimates from the U.S. export data. According to column (1), a megawatt-hour increase in per capita electricity consumption is associated with a 6.0% increase in the price of electric goods, where a megawatt-hour is approximately the difference in per capita electricity consumption between Zimbabwe and Turkey. This estimate is statistically significant at the 1% level of significance.

A typical concern in empirical work studying the determinants of export prices is that high prices reflect higher-quality goods. While the disaggregate nature of the data and the normalization of prices by their within-good standard deviation mitigate this concern to some extent, there may still remain scope for quality variation within an HS-10 category. To address this concern, Subsample 2 in Table I.2 drops from the sample all electric goods with long quality ladders. Specifically, I use the quality ladder estimates from Khandelwal (2010), and I drop all electric goods with ladder estimates above the median estimate.[20] The sample retains other consumer goods with long ladder estimates. Therefore, the regression will, if anything, understate the dependence of prices of electric goods on electricity access. This is because, to the extent that high export prices reflect high quality consumer goods sold to rich countries, the regression will estimate a high value of the country fixed effect for rich countries.[21] Since

---

[19] The qualitative results below are generally robust to omitting quantity from the regression.

[20] Approximately half of the HS-10 categories have nonmissing ladder estimates. Those with missing ladder estimates are kept in the sample. Note that long quality ladder estimates for a final good may reflect strong complementarity with catalyst goods, rather than high quality. This is because the estimates of ladder length in Khandelwal (2010) are based on the assumption that high market share (conditional on price) reflects high quality. In the models above, goods with high degrees of complementarity also have high market share. Thus dropping goods with long estimated quality ladders may remove some goods that are strong complements with catalyst goods, thus biasing downward the estimated relationship between catalysts and the prices of final goods.

[21] For a model predicting a relationship between quality of imports and income, see Hallak (2006).

electricity consumption is positively correlated with per capita income, some of the dependence of prices on housing will be captured by the high fixed effect estimates in rich countries. The estimates in column (2) are similar to those in column (1), suggesting that the results are driven by pricing-to-market rather than by quality differences.

The results in columns (1) and (2) of Table I.2 strongly support the hypothesis that prices of electric goods across countries depend on electricity access. However, the correlation between GDP per capita and MwH per capita for the sample of destination countries is 0.75 and the possibility remains that the estimate of β captures the dependence of prices of electric goods on a component of income that is not fully captured by the country level fixed effects. In other words, it is possible that electric good prices have an above average dependence on income per capita, and that the estimate of β is capturing this dependence. To verify that this positive estimate of β is driven by electricity access as a catalyst, rather than by other mechanisms associated with high incomes, column (3) reports the results from a modified version of specification (45) in which electricity consumption is interacted with log GDP per capita:

$$p_{ch} = \alpha_c + \gamma_h + \psi q_{ch} + \beta \text{MWHpercap}_c \text{Egood}_h + \beta_2 \text{GDPpercap}_c \text{Egood}_h + \epsilon_{ch}. \quad (46)$$

β_2 captures the extent to which electric goods are associated with high incomes per capita, conditional on country-specific determinants of consumer goods prices and conditional on the dependence of prices of electric goods on electricity access. According to column (3), the estimate of β_2 is not significantly different from zero, while the new estimate of β is lower and less significant. These results suggest that specification (46) lacks the power to distinguish the relative importance of electricity consumption per capita and income on the prices of electric goods. As we will see, the empirical test using Chinese export data is more powerful and indicates that there is a statistically and economically significant dependence of prices of electric goods on electricity access, even when conditioning on income per capita.

The results from the Chinese export data are qualitatively similar to the results from U.S. export data. Table I.3 shows that a MwH per capita increase in electricity consumption is associated with a statistically significant 2% to 3% increase in prices of electric goods. To the extent that the product-firm-firm×location dummies effectively condition on quality, the positive estimate of β from the Chinese data can be interpreted as evidence of pricing-to-market. Column (2) of Table I.3 shows that the estimate of β_2 is negative and insignificant, while the estimate of β is large and strongly significant. This suggests that any dependence of electric good prices on income is similar to the dependence of consumer goods prices on income as captured by the country-level fixed effects. Thus, the U.S. and Chinese export data appear to support the hypothesis that electricity access is a catalyst for demand for electric goods, and that electric goods are more expensive in countries with superior access to electricity.[22]

6.2 Housing Volume and Prices of Household Goods

Next, I assess whether prices of exports of households goods depend on European countries' stock of housing. Europe is an especially suitable region for such an investigation because its countries have low levels of within-country inequality, mitigating potential concerns that housing volume of the average resident may differ from housing volume of the consumer driving demand for household goods. Furthermore, housing volume is generally high in Europe, so a marginal increase in volume, such as an additional room, is likely to increase demand for furnishings of those rooms.[23]

The empirical specification is

$$p_{ch} = \alpha_c + \gamma_h + \psi q_{ch} + \beta \text{Vol}_c \text{HHgood}_h + \epsilon_{ch}, \qquad (47)$$

where Vol_c is the measure of the housing stock in country c, HHgood_h indicates whether the good is classified as a household good, and ϵ_{ch} denotes the regression error. The remaining variables are defined as above. All data are 2005 values, the only year for which data on Europe's housing stock is available. The baseline sample excludes all HS-10 products sold to less than 10 countries, and all product-country pairs for which less than 100 units were sold.

The coefficient β captures the extent to which the markup for household goods depends on housing volume. According to Column 1 in Table I.4, a standard deviation increase in a European country's housing volume index is associated with a 5.7% increase in the price of household goods. This estimate is significant at the 1% level of significance and is robust to dropping household goods with long quality ladders from the sample (column 2). This suggests

---

[22] Falsification exercises verify that other subsets of consumer goods (e.g. clothing, battery-powered goods, luxury goods, etc) do not have an above-average dependence on electricity consumption, which suggests that the positive dependence of electric goods prices on electricity consumption is indeed due to demand complementarity.

[23] In less developed regions, differences in volume are less likely to translate into marginal increases in demand for household goods; rather, in less developed countries, higher volume may imply an increase in personal space but not an increase in demand for furnishing.

that the estimated relationship between prices of household goods and a country's housing stock does *not* reflect high quality consumer goods being sold to countries with high housing volumes. Rather, the relationship reflects primarily a failure of the law of one price for household goods such that identical household goods are more expensive in countries with more housing per capita.

Column 3 shows the results from a modified version of equation (47) in which the interaction between log GDP per capita and an indicator for household goods is included as a regressor:

$$p_{ch} = \alpha_c + \gamma_h + \psi q_{ch} + \beta \mathrm{Vol}_c \mathrm{HHgood}_h + \beta_2 \mathrm{GDPpercap}_c \mathrm{HHgood}_h + \epsilon_{ch}. \quad (48)$$

The estimate of β_2 is not significantly different from zero, and the estimate of β remains large and significant, suggesting that housing is a catalyst that is associated with high prices of household goods and that the dependence of household goods prices on income is captured by the country-level fixed effects.

To determine which goods are driving this strong relationship, I reclassify goods into subcategories of household goods (e.g. dishwashers, kitchen appliances, etc.), and rerun the Subsample 2 regression by interacting housing volume with each subcategory. Television-related goods (e.g. antennas and satellite dishes) and refrigerators are the most important contributors to the observed relationship between a country's housing stock and the price it pays for household goods, followed closely by household furnishings. This result does not imply that housing does not complement demand for other household goods; rather, it is a reflection of the relatively high quantity of U.S. exports of television and refrigerator-related goods. Housing may complement demand for dishwashers, but U.S. exports of dishwashers to Europe are insufficient to provide a precise estimate of this relationship.

The results from specification (47) on Chinese export data correspond to those from the U.S. data. According to Column 1 in Table I.5, a standard deviation increase in housing volume is associated with a 1.6% increase in the prices of Chinese exports of household goods. However, the dependence of household goods prices on housing disappears under regression (48) in which $\mathrm{GDPpercap}_c \mathrm{HHgood}_h$ is included as a regressor. This may be due to the high correlation between GDP and housing across European countries in the sample (0.9), or it may be a consequence of the way in which housing volume is calculated. The ICP's measure of housing volume includes the service flow from the quality of the house (age of the house, heating quality, etc). If housing volume is more important than housing quality for demand for household goods, then the ICP measure will misrepresent the amount of housing catalyst across countries.

The ICP provides an alternative measure of housing volume based on the Consumption Equivalent Method (CEM), which assumes that housing volume is proportional to private consumption expenditures. The reader is referred to Heston (2011) for a more detailed comparison of the two measures. The two measures are very different for some countries, reflecting in part differences in the different weights placed on housing quality. Columns (3) and (4) of Table I.5 show that the dependence of prices of household goods on the CEM measure of housing volume is much higher than is predicted by the baseline measure.

Which housing measure is a more accurate measure of housing as a catalyst? One way to distinguish between the two measures is to see which predicts a higher dependence of prices of luxury goods on income per capita. Luxury goods are assumed to have an above-average dependence on income per capita, and an ambiguous (but likely average) dependence on housing. I identify luxury goods as those related to water sports, tennis, golf, skiing, and adventure sports. According to Columns (5) and (6), the baseline housing measure predicts an above-average dependence of luxury prices on housing, while the CEM measure predicts an above-average dependence of luxury prices on income per capita. By this criterion, therefore, it appears that the alternative CEM housing measure is the more accurate measure of housing volume, and that prices of household goods have an above-average dependence on housing volume.

Using the CEM measure in place of the baseline volume measure on the U.S. data is less conclusive. The coefficients on the CEM measure and on GDP per capita are both positive but statistically insignificant (not shown). The dependence of luxury goods prices on the CEM measure is negative but insignificant. Thus it's not clear that the CEM measure more accurately captures the aspect of housing that is the relevant catalyst for household goods produced in the U.S.

It is possible that the CEM measure, which may be a better measure of volume, more accurately captures the catalyst for Chinese-produced household goods, while the baseline measure, which is perhaps a better measure of quality, more accurately captures the catalyst for U.S.-produced household goods. This would be the case if, for example, housing quality is a catalyst for higher-quality household goods, and the U.S. produces higher-quality household goods than does China.

6.3 Paved Roads and Prices of New Cars

Data on the percent of paved roads are available across regions for different years between 2003 and 2006. I take the most recent year for which data are available in a country as that country's measure of road quality and estimate the following specification:

$$p_{ch} = \alpha_c + \gamma_h + \psi q_{ch} + \beta \mathrm{Road_c Newcar_h} + \epsilon_{ch}, \qquad (49)$$

where $\mathrm{Road_c}$ is the percent of roads that are paved in country c and $\mathrm{Newcar_h}$ indicates whether good h is a new car. The remaining variables are defined as above. Specification (49) is tested only on U.S. data since the Chinese Customs data do not include sales of new cars in 2005. Unit values and quantities for each country-product pair are averages of the values between 2004 and 2006 (the three most recent years of data). The sample excludes all HS-10 products sold to less than 10 countries, and all product-country pairs for which less than \$10,000 worth of goods were sold. [24]

Table I.6 shows the results from specification (49). In columns (1) and (2), the sample includes all exported non-military goods (end use classification 0 through 4). Column (1) states that a percentage point increase in the fraction of roads that are paved is associated with a 0.6 percent increase in the price of new cars. This relationship is statistically significant. While prices of new cars depend on road quality, column (2) suggests that the paved roads are not associated with high prices of other automobiles or auto parts. To corroborate the evidence in column (2), columns (3) and (4) restrict the sample to auto-related exports so that country-level fixed effects are determined by the relationship of prices of auto-related goods across countries. Consistent with the evidence in regression (1), regressions (3) and (4) show that road quality is associated with high prices of new cars, even conditional on country-level determinants of prices of auto-related goods. The relationship between road quality and prices of new cars is statistically significant and is robust to the inclusion of log GDP per capita interacted with an

---

[24] The sample does not restrict observations based on the quantity of goods because cars are assumed to be sold in lower quantities on average than are consumer goods. Indeed, using the same cutoff threshold of 100 units in Sections 6.1 and 6.2 would remove almost two-thirds of the new car observations from the sample.

indicator variable for new cars as a regressor, suggesting that road quality is associated with high prices of new cars conditional on any price association due to destination-country income.[25]

The evidence in Table I.6 suggests that road quality complements demand for new cars but not demand for automobiles generally and auto parts. One possible explanation for this result is that demand for automobiles (used or new) is driven primarily by the need for transportation, regardless of the quality of the roads. Demand for new cars relative to used cars, however, depends on the enjoyment of driving, in addition to efficient travel. A new Cadillac is not much more effective than an old jeep at transporting an individual over a mile of dirt road. However, a luxury Cadillac may be more effective at transporting someone on paved roads, and it is likely to be a more comfortable experience.

An additional explanation for the insignificant relationship between auto parts in general and paved roads is that demand for auto parts may be high when roads are in poor condition. Not only are consumers less likely to purchase new cars (for which new parts are not immediately necessary) when roads are poor, but bad roads cause car damage and thus necessitate constant repair and frequent need for replacement parts.

The general message from Table I.6 is consistent with the model's predictions based on demand complementarity and pricing-to-market: road quality is associated with higher prices of new cars. The main caveat is that the results may simply reflect the fact that higher quality cars are sold to countries with higher quality roads. Since new cars have long quality ladders, this concern cannot be addressed as in Sections 6.1 and 6.2 by dropping products with short quality ladders.

In reality, high prices of new cars are likely a result of both sales of high-quality cars and pricing to market for identical models. The standard assumption in the literature has been that price differences reflect quality differences, but recent evidence has demonstrated a strong role for price discrimination across countries for a range of products.[26] Thus it seems reasonable to infer price discrimination in the auto market as well. Precisely identifying the relative importance of pricing-to-market in the auto industry will require price data on identical models.

---

[25] It is not surprising that the estimate of β remains significant even with the inclusion of GDP on the right-hand-side of (53) since GDP per capita and road quality have a relatively low correlation across countries of 0.58.

[26] See, for example, Alessandria and Kaboski (2011) for evidence across a range of goods and countries and Simonovska (2011) for evidence across a specific category of goods within Europe.

7. Discussion of Catalyst Goods

According to the empirical results, catalyst consumption is associated with higher prices of relevant tradables. One advantage of the empirical specification is that the strong estimated relationship is conditional on the association between catalyst consumption and prices that is captured by the country-level fixed effects, and thus provides lower bound on the dependence of prices on catalyst goods. Since catalyst consumption is strongly correlated with income per capita, the results also provide a lower bound on the dependence of consumer prices on income per capita driven by demand complementarities. A limitation of this approach is that I can neither rule out other mechanisms nor quantify their roles because the country fixed effects capture the average dependence of prices on income per capita without distinguishing precise mechanisms. Thus it seems reasonable to infer that demand complementarities are an important source of price variation, but it is left for future work to determine precisely how important relative to other causes of the price-income relationship.

The catalyst goods examined in the empirical section are durables such as housing and public infrastructure. These goods are readily identified as catalysts for specific subsets of tradable goods and are thus amenable to an empirical investigation of the role of catalysts in generating high consumer goods prices in rich countries. However, the notion of a catalyst applies broadly to any good or service that may complement demand for other goods and services. This includes nondurable goods and services, as well as amenities for which there is not an explicit market price. Customers do not always directly pay for amenities associated with the services they purchase, but the amenities complement their demand for services. For example, customers may have higher utility from food at a restaurant if the restaurant has nice artwork, good service, and comfortable chairs. The more efficiently a restaurant can produce these complementary goods and services, the more it can charge for food of a given cost. Likewise, the availability of retail stores, and the quality of service at those stores, can complement demand for retail goods. Demand complementarities at the retail level can explain, for example, the finding in Crucini, Telmer, and Zachariadis (2005) that nontraded retail inputs account for much of the price dispersion for goods and services in the European Union.

Finally, one can think of marketing and related sales activity as catalyst services. A number of recent papers, including Arkolakis (2010) and Gourio and Rudanko (2011), investigate the implications of marketing and sales activity on firm outcomes. The analysis above suggests that such activity may also contribute to the cross-country differences in prices and real investment rates if such activity increases consumers' demand.

8. Conclusion

A well-established empirical regularity is that tradable consumption goods are more expensive in countries with high per-capita incomes. This paper proposes a simple explanation for this relationship based on demand complementarities and pricing-to-market by monopolistically competitive firms: The utility consumers derive from tradable goods depends on their consumption of complementary goods. Rich countries can afford more complementary goods, which generates high (and inelastic) demand for tradables. As a result, monopolistically competitive firms charge higher markups in rich countries.

The paper provides direct empirical evidence that the phenomenon of demand complementarities and pricing-to-market is responsible for high prices of specific subsets of tradable goods in countries with high consumption of relevant complementary goods, conditional on income per capita and on other destination country-level determinants of prices. Specifically, household goods are sold at higher prices to countries with more housing volume per capita; electronic goods are sold at higher prices to countries with superior electricity infrastructure (as proxied by electricity consumption per capita); and new cars are sold at higher prices to countries with a higher percentage of paved roads.

The theoretical models developed in the paper demonstrate that evidence of demand complementarity and pricing-to-market also strongly supports the notion that nontradable consumer goods are more expensive in rich countries because demand is higher (and less elastic) there, thus offering an explanation to a longstanding puzzle in the trade literature. In addition, the evidence lends support to the notion that real investment rates are higher in rich countries due to high demand arising from higher wage-to-rental ratios in rich countries. Understanding why rich countries have higher rates of investment is important for understanding why income disparities persist between rich and poor countries. Economists have typically attributed differences in investment rates to market distortions (e.g. high taxes and corruption) in poor countries. Hsieh and Klenow (2007) argue that distortionary taxes cannot account for the relationship between investment rates and income per capita and instead prefer an explanation based on sectoral productivity differences. While much work remains to be done to quantify the precise role of demand complementarities in accounting for differences in prices and real investment rates across countries, my results suggest that low investment rates in poor countries may be due to low consumption of complementary goods and services rather than to distortionary taxes or within-country productivity differentials.

Appendix I.A

The models presented above feature an endowed numeraire that enters the utility function linearly. This setup is chosen for its tractability and because it permits a focus on demand complementarities, rather than the marginal utility of income, as the determinant of consumers' price elasticity of demand for final goods. Here I present an alternative closed-economy setup in which the numeraire is produced by labor, rather than endowed. The utility function is also altered to permit the marginal utility of income to vary with income.[27]

The representative agent's utility function is defined over the catalyst C, the mass Ω of final goods, and a numeraire Y:

$$U = Y^{\eta}\left(C^{\alpha}\int_{\Omega} f_{\omega}d\omega - \frac{1}{2}\gamma\int_{\Omega} f_{\omega}^{2}d\omega\right)^{1-\eta}, \tag{50}$$

where f_{ω} is consumption of final good $\omega \in \Omega$. This utility function is similar to that in Chaney (2008) in that it features Cobb-Douglass preferences over a homogenous numeraire and differentiated consumer goods.

The budget constraint is

$$wL + \int_{\Omega}\Pi_{\omega}d\omega = Y + p_C C + \int_{\Omega} p_{\omega}f_{\omega}d\omega, \tag{51}$$

Consumer optimization with respect to f_{ω} yields the implicit demand for final good of variety ω:

$$Y^{\eta}\,(1-\eta)B^{-\eta}(C^{\alpha} - \gamma f_{\omega}) = \lambda p_{\omega},$$

where λ is the multiplier on the budget constraint (51) and $B \equiv C^{\alpha}\int_0^{\Omega} f_{\omega}d\omega - \frac{1}{2}\gamma\int_0^{\Omega} f_{\omega}^{2}d\omega$ is the bundle of final and catalyst goods. We can obtain an expression for λ from the first order condition with respect to Y:

$$\eta Y^{\eta-1}B^{1-\eta} = \lambda.$$

Combining the above two equations yields an explicit expression for demand for final good ω:

$$f_{\omega} = \frac{1}{\gamma}\left[C^{\alpha} - \frac{\eta}{1-\eta}\frac{1}{Y}Bp_{\omega}\right]. \tag{52}$$

---

[27] If the model were to feature a numeraire produced by labor and a baseline utility function given by (1), the model solution would be at a corner in which the numeraire is the only good produced and consumed. A derivation of the corner solution to this alternative setup is available upon request.

Production of final goods, catalyst goods, and the numeraire good are linear in labor using labor productivity A, which is assumed to be identical across sectors. The final goods sector is monopolistically competitive, while the catalyst and numeraire sectors are perfectly competitive. Firm ω maximizes $\Pi_\omega = \left(p_\omega - \frac{w}{A}\right) f_\omega$, which implies the optimal price

$$p_\omega = \frac{1}{2}\left[Y\frac{1-\eta}{\eta}B^{-1}C^\alpha + \frac{w}{A}\right]. \tag{53}$$

The price increases with C^α, as in Section 2. It also increases as the marginal utility of income falls. Since λ is decreasing in Y, the price of final goods is increasing in Y.

Given the price the resulting demand for good ω is

$$f_\omega^d = \frac{1}{2\gamma}\left[C^\alpha - \frac{w}{A}\frac{\eta}{1-\eta}\frac{1}{Y}B\right]. \tag{54}$$

Demand for the catalyst is derived from consumer optimization:

$$C = \left[Y\frac{1-\eta}{\eta}\alpha\frac{F}{p_C}B^{-1}\right]^{\frac{1}{1-\alpha}}. \tag{55}$$

Equilibrium is characterized by demand for catalysts, demand for consumer goods, and labor market clearing,

$$L = \frac{1}{A}(\Omega f + C + Y) \tag{56}$$

These conditions can be written as

$$C = \left[Y\frac{1-\eta}{\eta}\alpha f\left(C^\alpha f - \frac{1}{2}\gamma f^2\right)^{-1}\right]^{\frac{1}{1-\alpha}}$$

$$f = \frac{1}{2\gamma}\left[C^\alpha - \frac{\eta}{1-\eta}\frac{\Omega}{Y}\left(C^\alpha f - \frac{1}{2}\gamma f^2\right)\right]$$

$$L = \frac{1}{A}(\Omega f + C + Y),$$

where I've substituted in $w = A$ and $p_C = w/A$. Figure I.A1 shows market responses to an increase in productivity A. As in the baseline model in Section 2, prices of final goods are increasing in a country's wealth due to markups that increase with consumption of the catalyst good.

Appendix I.B

The models in this paper use a simple linear demand curve to illustrate how an increase in complementary goods (catalysts) reduces the price-elasticity of demand for final consumer goods by shifting out the demand curve. Linearity of the demand curve is sufficient for a decrease in the price elasticity of demand in response to an increase in the complementary good, but it is not a necessary condition. This appendix derives the necessary and sufficient conditions on the demand curve under which an increase in complementary goods leads to higher markups for consumer goods.

A generic demand curve can be written $q = q(C, p)$, where C is the complementary catalyst and p is the price of the good. The price-elasticity of demand is decreasing in C if and only if $\frac{\partial \epsilon}{\partial C} < 0$, where $\epsilon \equiv \left|\frac{\partial q}{\partial p}\frac{p}{q}\right|$. We can write $\frac{\partial \epsilon}{\partial C} = -q_{21}\frac{p}{q(D,p)} + q_2\frac{p}{q(D,p)^2}q_1$, in which case the necessary and sufficient condition simplifies to

$$qq_{21} > q_2q_1. \tag{57}$$

Condition (57) states that any slope-increasing effects of an increase in C on the demand curve must be more than compensated by a shift out of the demand curve. In the commonly used case of a constant elasticity demand curve, $q = Cp^{-\epsilon}$, these two effects exactly cancel out so that $qq_{21} = q_2q_1$. As discussed in Nakamura and Zerom (2010), price-independent demand elasticities are difficult to reconcile with the data. Their estimates on coffee demand suggest that the price elasticity of demand is increasing in the price.

Appendix I.C

This appendix alters the model in Section 6 by assuming that differentiated investments goods are produced under monopolistic competition and aggregated into a final investment good through a CES aggregator. Each country $j \in \{N, S\}$ produces a mass Ψ_{j} of differentiated investment goods. Each good $\psi_j \in \Psi_{\mathrm{j}}$ is exported and sold domestically. Countries N and S purchase investment goods and costlessly aggregate them into a final investment good.

Equation (42) changes to

$$I_j = \left(\sum_{i=N,S} \int_{\psi_i \in \Psi_{\mathrm{j}}} q_j(\psi_i)^{\frac{\sigma-1}{\sigma}} d\omega_i \right)^{\frac{\sigma}{\sigma-1}}, \tag{58}$$

where $q_j(\psi_i)$ is country j's quantity of the differentiated intermediate investment variety ψ_i produced in country i. Demand for good ψ_i in country j is

$$q_j(\psi_i) = \left(\frac{p_{q_j}(\psi_i)}{P_{Ij}} \right)^{-\sigma} I_j,$$

And the optimal price charged by firm ψ_i in country j is

$$p_{q_j}(\psi_i) = \frac{\sigma}{\sigma-1} c_i. \tag{59}$$

Note that the price is a constant markup over marginal costs, so each differentiated investment good is sold at the same price in both countries. Therefore the cost of final investment goods equalizes across countries, as does the rental rate of capital. As in Section 6, the wage is higher in the rich country (N), which causes higher demand for capital in N.

The equilibrium conditions are altered only slightly relative to those in Section 6. The trade balance condition now accounts for the fact that both countries produce investment goods:

$$y_N^0 - y_N + \Psi_{\mathrm{N}} p_{qSN} q_{SN} + \Omega_N p_{SN} f_{SN} = \Omega_S p_{NS} f_{NS} + \psi_{\mathrm{N}} p_{qNS} q_{NS}, \tag{60}$$

where p_{qij} and q_{ij} are defined analogously to p_{ij} and f_{ij} for $i, j \in \{N, S\}$. Demand for labor also now accounts for investment good production in both countries:

$$L_N = \left(\frac{R}{w_N} \frac{\eta}{1-\eta} \right)^{1-\eta} \left[\frac{\Omega_N (f_{NN} + f_{SN})}{A_N} + \delta \frac{C_N}{A_{CN}} + \frac{q_{NN} + q_{SN}}{A_{IN}} \right],$$

$$L_S = \left(\frac{R}{w_S} \frac{\eta}{1-\eta} \right)^{1-\eta} \left[\frac{\Omega_S (f_{SS} + f_{NS})}{A_S} + \delta \frac{C_S}{A_{CS}} + \frac{q_{SS} + q_{NS}}{A_{IN}} \right].$$

Figure I.C1 shows how relative final goods prices, real investment, and investment prices depend on wealth in N. As in Section 6, the patterns of prices and investment are consistent with facts (1 through 3).

Appendix I.D

This appendix departs from the model in Section 5 by postulating that final investment goods are produced from differentiated investment goods using a quadratic aggregator similar to the utility function, thus permitting price-dependent markups for differentiated investment goods. Specifically, equation (58) is now

$$I_j = L_{Ij} + \sum_{i=N,S} \int_{\psi_i \in \Psi_i} \left[\theta q_j(\psi_i) - \frac{\gamma}{2} \left(q_j(\psi_i) \right)^2 \right] d\omega_i, \tag{61}$$

where L_{Ij} is labor in country j that is allocated to the aggregation of investment goods. Before proceeding, a couple of remarks must be made regarding this particular aggregator function. First, there is a bliss point after which additional units of a given differentiated investment good are actually counterproductive. In the utility function, the bliss point represents the fact that more consumer goods eventually becomes undesirable (consider eating a hundred cheeseburgers in a day). It is less clear what a bliss point represents in the aggregation of investment goods. Thus the CES aggregator may actually be more realistic than the quadratic investment aggregator. Nonetheless, this section presents the quadratic aggregator for completeness.

Second, equation (61) features labor as a substitute for intermediate investment goods. This assumption captures the notion that with sufficient labor input, the aggregate investment good could be produced without any intermediate investment goods. It also simplifies the analysis.

Demand in country j for investment good ψ_i is

$$q_j(\psi_i) = \frac{1}{\gamma} \left[\theta - \frac{1}{w_j} p_{q_j}(\psi_i) \right]$$

The optimal price charged by firm ψ_i for a good sold to country j is

$$p_{q_j}(\psi_i) = \frac{1}{2} \left(\theta w_j + c_i \right),$$

And resulting demand is

$$q_j(\psi_i) = \frac{1}{2\gamma} \left[\theta - \frac{c_i}{w_j} \right].$$

The price of the final investment good in country j is equal to the wage in j, and investment demand in j is proportional to demand for capital in j.

In contrast to the prior models with investment, the price of investment does not equalize across countries; nor does the rental price of capital.[28] To determine relative prices, we must solve for the equilibrium. There are 20 unknowns, $w_N, w_S, y_N, y_S, C_N, C_S, f_{NN}, f_{NS}, f_{SS}, f_{SN}, p_{NS}, p_{SN}, R_N, R_S, K_N, K_S, q_{NN}, q_{NS}, q_{SS}$, and q_{SN}, for which I solve using the following equilibrium equations:

$$C_N = \left[\frac{\alpha A_{CN}(\Omega_N f_{NN} + \Omega_S f_{NS})}{c_N(r+\delta)}\right]^{\frac{1}{1-\alpha}} \qquad C_S = \left[\frac{\alpha A_{CS}(\Omega_S f_{SS} + \Omega_N f_{SN})}{c_S(r+\delta)}\right]^{\frac{1}{1-\alpha}}$$

$$L_N = \left(\frac{R_N}{w_N}\frac{\eta}{1-\eta}\right)^{1-\eta}\left[\frac{\Omega_N(f_{NN}+f_{SN})}{A_N} + \delta\frac{C_N}{A_{CN}} + \frac{q_{NN}+q_{SN}}{A_{IN}} + L_{IN}\right].$$

$$L_S = \left(\frac{R_S}{w_S}\frac{\eta}{1-\eta}\right)^{1-\eta}\left[\frac{\Omega_S(f_{SS}+f_{NS})}{A_S} + \delta\frac{C_S}{A_{CS}} + \frac{q_{SS}+q_{NS}}{A_{IN}} + L_{IS}\right]$$

$$y_N^0 - y_N + \Psi_N p_{qSN} q_{SN} + \Omega_N p_{SN} f_{SN} = \Omega_S p_{NS} f_{NS} + \Psi_S p_{qNS} q_{NS}$$

$$y_N^0 + y_S^0 = y_N + y_S$$

$$R_N = p_{IN}(r+\delta) \qquad R_S = p_{IS}(r+\delta)$$

$$f_{NN}^c = \frac{1}{2\gamma}\left(C_N^\alpha - \frac{c_N}{A_N}\right) \quad f_{SN}^c = \frac{1}{2\gamma}\left(C_S^\alpha - \frac{c_N}{A_N}\right)$$

$$f_{SS}^c = \frac{1}{2\gamma}\left(C_S^\alpha - \frac{c_S}{A_S}\right) \quad f_{NS}^c = \frac{1}{2\gamma}\left(C_N^\alpha - \frac{c_S}{A_S}\right)$$

$$p_{NS} = \frac{1}{2}\left(C_N^\alpha + \frac{c_S}{A_S}\right) \quad p_{SN} = \frac{1}{2}\left(C_S^\alpha + \frac{c_N}{A_N}\right),$$

$$K_N = \frac{w_N}{R}\frac{1-\eta}{\eta}L_N \qquad K_S = \frac{w_S}{R}\frac{1-\eta}{\eta}L_S$$

---

[28] The differences in rental rates across countries hinges on the implicit assumption that markets for capital assets are separate across countries. Permitting cross-country capital ownership would cause rental rates and investment prices to equalize across countries, thus defeating the purpose of the exercise in this section of demonstrating that simple model extensions can generate a positive relationship between investment prices and income per capita. As discussed in Hsieh and Klenow (2007), the evidence of a positive relationship between investment prices and income per capita is limited primarily to prices of investment structures, which are highly nontraded. Thus the models in Section 5 and Appendix I.C are likely more empirically relevant than the model in this appendix.

$$q_{NN} = \frac{1}{2\gamma}\left[\theta - \frac{c_N}{w_N}\right] \qquad q_{SN} = \frac{1}{2\gamma}\left[\theta - \frac{c_N}{w_S}\right]$$

$$q_{SS} = \frac{1}{2\gamma}\left[\theta - \frac{c_S}{w_S}\right] \qquad q_{NS} = \frac{1}{2\gamma}\left[\theta - \frac{c_S}{w_N}\right]$$

where

$$p_{q_{NN}} = \frac{1}{2}[\theta w_N + c_N] \qquad p_{qNS} = \frac{1}{2}[\theta w_N + c_S]$$

$$p_{q_{SS}} = \frac{1}{2}[\theta w_S + c_S] \qquad p_{qSN} = \frac{1}{2}[\theta w_S + c_N]$$

$$c_N = \frac{w_N}{\eta} L_N \qquad c_S = \frac{w_S}{\eta} L_S$$

Figure I.D1 shows how market outcomes depend on productivity in country N. Relative prices in N are increasing in productivity in N, and purchases of intermediate investment goods are higher in N, consistent with the stylized facts discussed in Hsieh and Klenow (2007).

Table I.1: Consumer Goods and Household Goods by End Use

End Use code	End Use	Household Good?	
40000	Apparel, household goods - textile	\	Include towels, bed linens, curtains
40030	Apparel,household goods-nontextile	\	Include towels, bed linens, curtains
40050	Sports apparel and gear		
40100	Pharmaceutical preparations		
40110	Books, printed matter		
40120	Toiletries and cosmetics		
40130	Tobacco, manufactured		
40140	Writing and art supplies		
41000	Furniture, household goods, etc.	X	
41010	Glassware, chinaware	X	
41020	Cookware, cutlery, tools	X	
41030	Household appliances	\	Exclude Radiators, Air Conditioners
41040	Rugs	X	
41050	Other household goods	\	Exclude shavers, hair dryers, cellular phones
41110	Pleasure boats and motors		
41120	Toys/games/sporting goods		
41140	Musical instruments		
41200	TV's, VCR's, etc.	X	
41210	Stereo equipment, etc.	X	
41220	Records, tapes, and disks		
41300	Numismatic coins		

Note: The table shows consumer goods by end use classification. X indicates that all goods in an end use category are identified as household goods. / indicates that a subset of goods in that category are identified as household goods.

Table I.2-Coefficient Estimates from Fixed-Effects Regressions of Log Unit Values of U.S. Exports on PerCapita Electricity Consumption

Dependent Variable: Log(price)/SD(Log(price))

	Subsample 1	Subsample 2	
Regressors	(1)	(2)	(3)
MWh per capita X Electric good	0.060***	0.066***	0.041
	(0.018)	(0.022)	(0.028)
log(GDP per capita) X Electric good			0.045
			(0.035)
Log(quantity)/SD(Log(quantity))	-0.613***	-0.614***	-0.614***
	(0.010)	(0.011)	(0.011)
Product FEs	YES	YES	YES
Country FEs	YES	YES	YES
R-squared	0.29	0.29	0.29
# observations	24,061	23,632	23,632
# products	1,309	1,281	1,281

Notes: Prices and quantities are normalized by their within-product standard deviation. Data source: World Bank Development Indicators and U.S. Exports by HS-10 classification. Subsample 1 includes all consumer goods which are sold to at least 10 countries, and all product-country observations with at least 100 units sold. Subsample 2 drops from Subsample 1 all electric goods with quality ladder estimates greater than the median, where the quality ladder estimates are obtained from Khandelwal (2011). Robust standard errors clustered at the product level in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Table I.3-Coefficient Estimates from Fixed-Effects Regressions of Log Unit Values of Chinese exports on PerCapita Electricity Consumption

Dependent Variable: Log(price)/SD(Log(price))		
Regressors	(1)	(2)
MWh per capita X Electric good	0.027**	0.037**
	(0.011)	(0.015)
log(GDP per capita) X Electric good		-0.019
		(0.025)
Log(quantity)/SD(Log(quantity))	-0.261***	-0.261***
	(0.012)	(0.012)
Product Firm City Zip FEs	YES	YES
Country FEs	YES	YES
R-squared	0.06	0.06
# observations	158,400	158,168
# product-firm-firmXlocations	26,276	26,276

Notes: Prices and quantities are normalized by their within-product standard deviation. Data source: World Bank Development Indicators and Chinese Exports by HS-8 classification. Robust standard errors clustered at the product level in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Table I.4-Coefficient Estimates from Fixed-Effects Regressions of Log Unit Values of U.S. Exports on Housing Volume in European Countries

Dependent Variable: Log(price)/SD(Log(price))

	Subsample 1	Subsample 2	
Regressors	(1)	(2)	(3)
Housing volume X Household good	0.057***	0.056***	0.059*
	(0.022)	(0.024)	(0.033)
log(GDP per capita) X Household good			-0.008
			(0.067)
Log(quantity)/SD(Log(quantity))	-0.508***	-0.507***	-0.507***
	(0.016)	(0.017)	(0.017)
Product FEs	YES	YES	YES
Country FEs	YES	YES	YES
R-squared	0.18	0.17	0.17
# observations	9,646	9,125	9,125
# products	1,124	1,049	1,049

Notes: Prices and quantities are normalized by their within-product standard deviation. Data source: World Bank Development Indicators, and U.S. Exports by HS classification. Subsample 1 includes all consumer goods which are sold to at least 10 countries, and all product-country observations with at least 100 units sold. Subsample 2 drops from Subsample 1 all electric goods with quality ladder estimates greater than the median, where the quality ladder estimates are obtained from Khandelwal (2011). Robust standard errors clustered at the product level in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Table I.5-Coefficient Estimates from Fixed-Effects Regressions of Log Unit Values of Chinese Exports on Housing Volume in European Countries

Dependent Variable: Log(price)/SD(Log(price))						
Regressors	(1)	(2)	(3)	(4)	(5)	(6)
Housing volume X Household good	0.016*	-0.017				
	(0.009)	(0.014)				
Housing volume (alternative measure) X Household good			0.036***	0.027		
			(0.011)	(0.020)		
log(GDP per capita) X Household good		0.067***		0.015		
		(0.024)		(0.027)		
Housing volume X Luxury good					0.062**	
					(0.030)	
Housing volume (alternative measure) X Luxury good						-0.085
						(0.057)
log(GDP per capita) X Luxury good					0.113	0.284***
					(0.075)	(0.067)
Log(quantity)/SD(Log(quantity))	-0.237***	-0.237***	-0.237***	-0.237***	-0.236***	-0.236***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Product Firm City Zip FEs	YES	YES	YES	YES	YES	YES
Country FEs	YES	YES	YES	YES	YES	YES
R-squared	0.05	0.05	0.05	0.05	0.05	0.05
# observations	181,170	181,170	181,170	181,170	181,170	181,170
# product-firm-firmXlocations	27,250	27,250	27,250	27,250	27,250	27,250

Notes: Prices and quantities are normalized by their within-product standard deviation. Data source: World Bank Development Indicators, and Chinese Exports by HS-8 classification. Robust standard errors clustered at the product level in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Table I.6-Coefficient Estimates from Fixed-Effects Regressions of Log Unit Values of U.S. Exports on Percent of Paved Roads

Dependent Variable: Log(price)/SD(Log(price))	Sample: Auto Vehicles, Parts, and Engines	
Regressors	(1)	(2)
Percent of roads paved X New car	0.005***	0.003**
	(0.001)	(0.001)
log(GDP per capita) X New car		0.075
		(0.061)
Log(quantity)/SD(Log(quantity))	-0.544***	-0.546***
	(0.032)	(0.032)
Product FEs	YES	YES
Country FEs	YES	YES
R-squared	0.24	0.24
# observations	8,155	8,155
# products	136	136

Note: Prices and quantities are normalized by their within-product standard deviation. Data source: World Bank Development Indicators and U.S. Exports by HS-10 classification. The sample auto-related goods which are sold to at least 10 countries, and all product-country observations with at least $10,000 in value. Robust standard errors clustered at the product level in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Figure I.1: Comparative Statics: Market Outcomes as Productivity Increases.

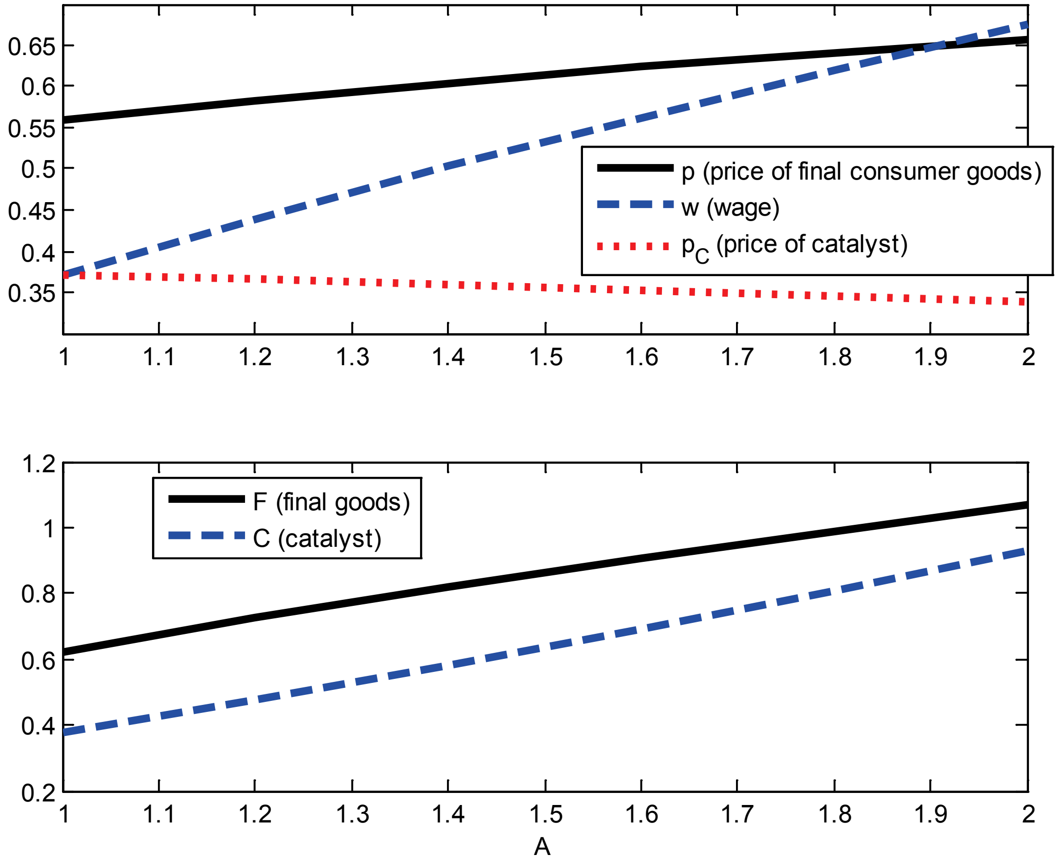
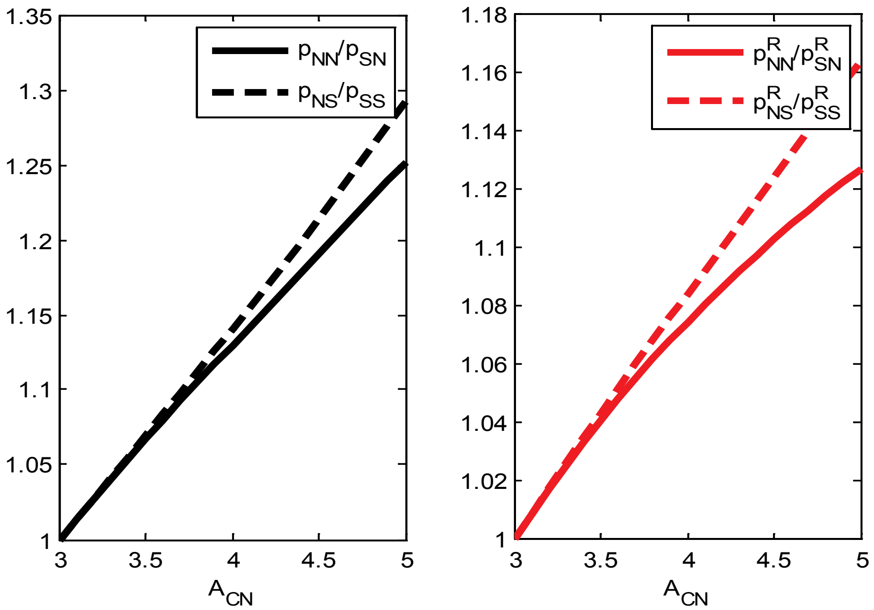


Figure I.2: Effect of Productivity in *N*’s Catalyst Sector on Relative Prices.



Note: The graph on the left shows the ratio of prices relative to the numeraire, while the graph on the right shows the ratio of PPP prices. See Footnote 9 for an explanation of how PPP prices are computed.

Figure I.3: Effect of Productivity in *N*'s Final Good Sector on Prices and Quantities.

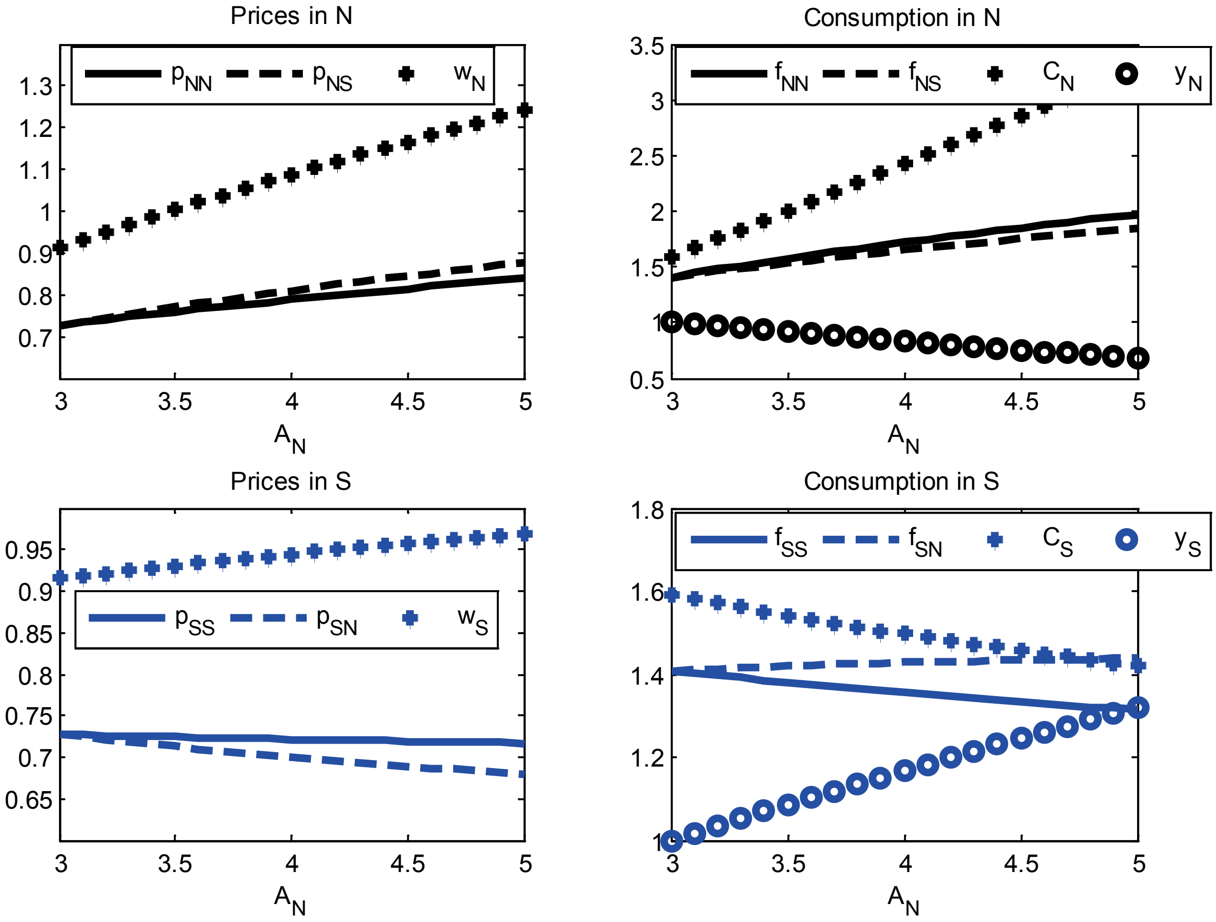


Figure I.4: Effect of Productivity in *N* on Welfare.

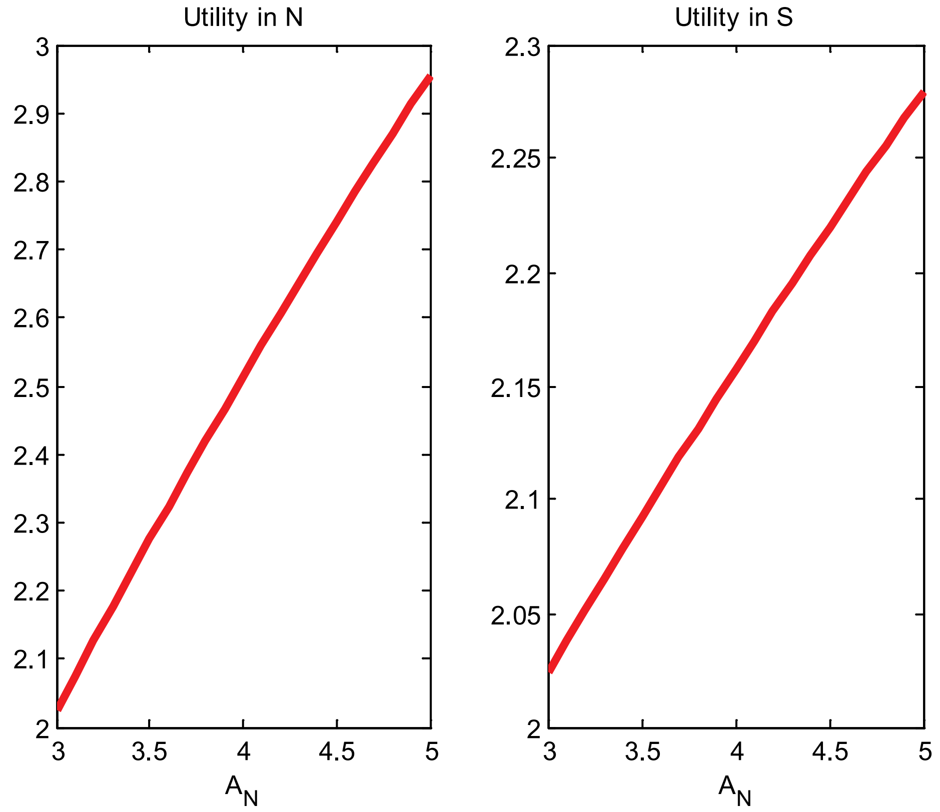


Figure I.5: Effect of Productivity in *N* on Market Outcomes in Two-Country Model with Nontraded Goods and Services.

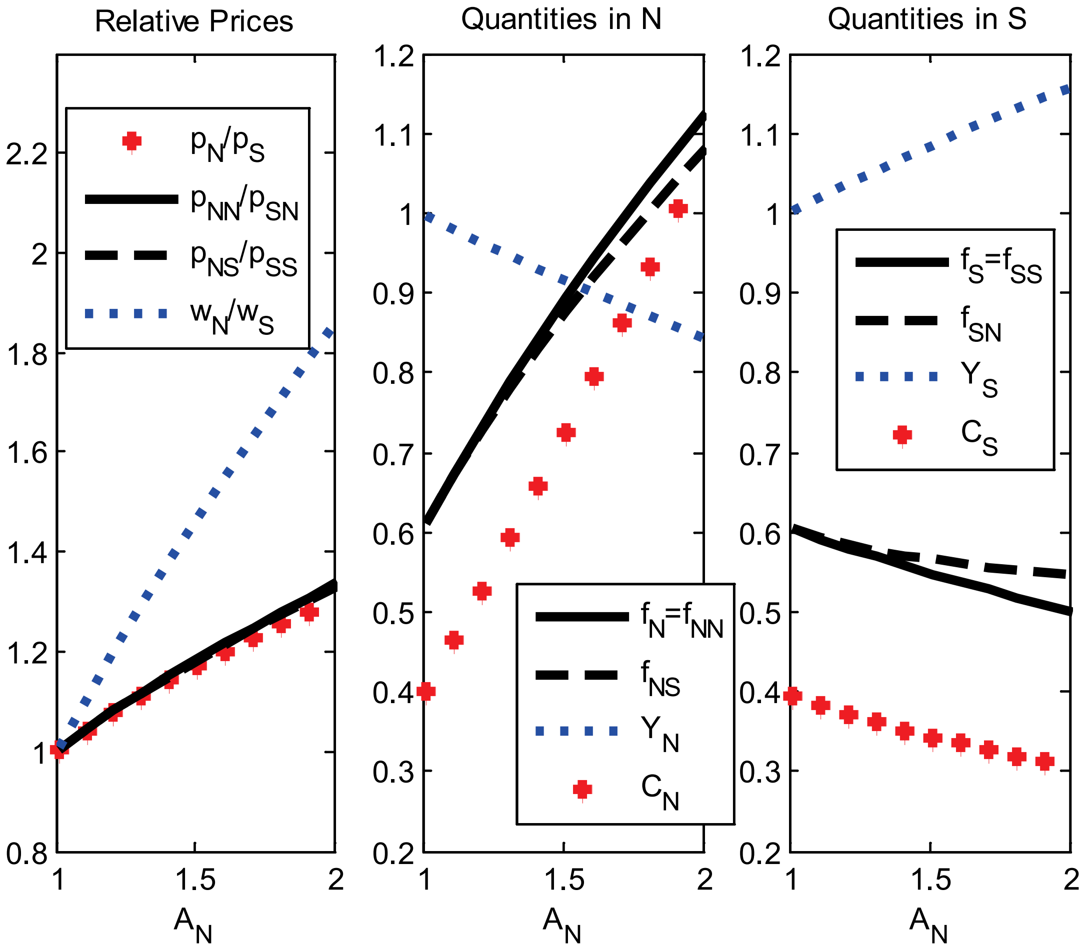
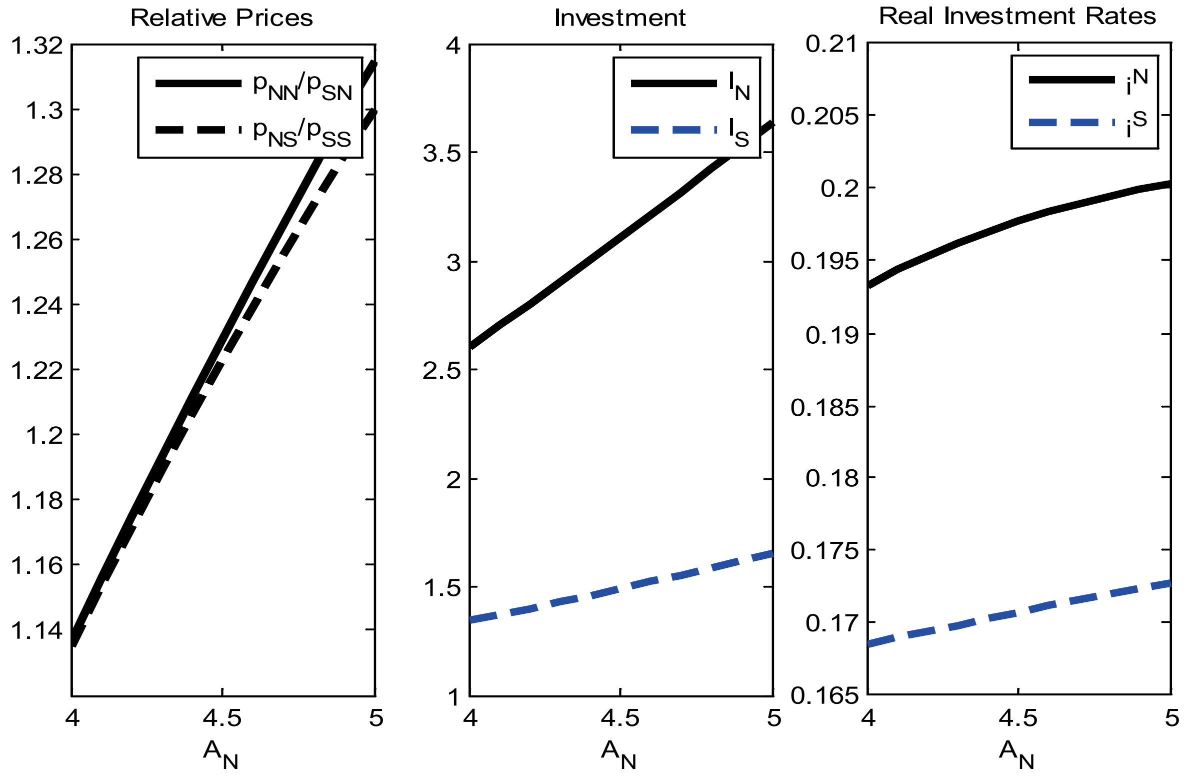


Figure I.6: Relative Prices and Real Investment in Two-Country Model with Capital.



Note: The real investment rate in country *j* is given by

$$i^j = \frac{p_I I_j}{y_j + \Omega_N p_{NN} f_{jN} + \Omega_S p_{NS} f_{jS} + p_{Xj} X_j + p_I I_j}$$

Figure I.A.1: Market Outcomes as Productivity Increases in Model with Numeraire Produced by Labor.

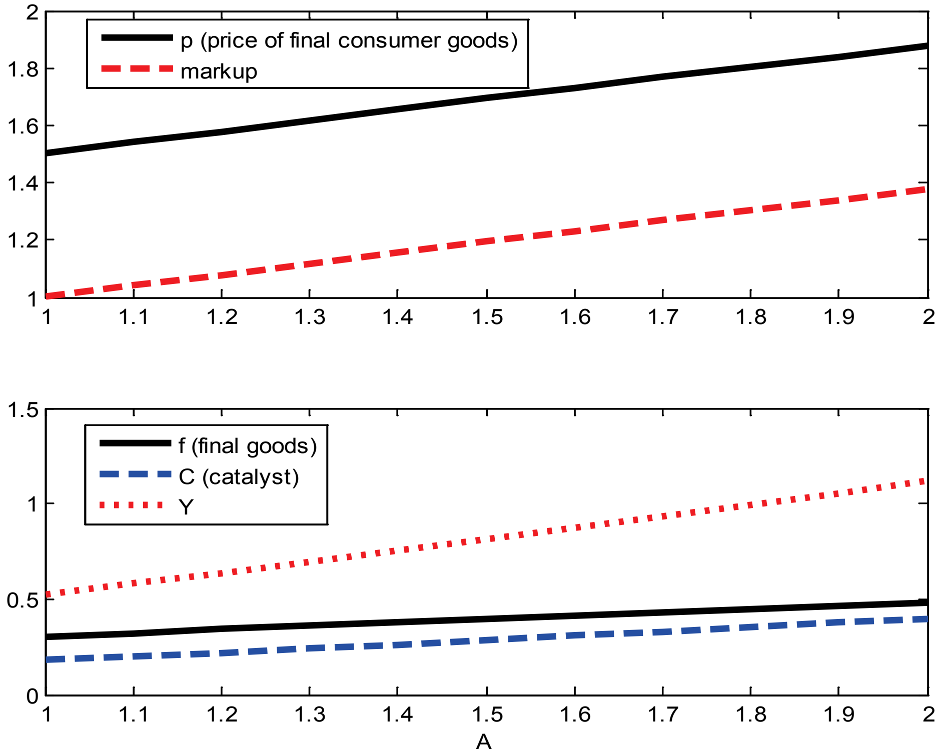


Figure I.C.1: Relative Prices and Real Investment in Alternative Model with Investment Produced by a CES Aggregator over Differentiated Intermediate Investment Goods.

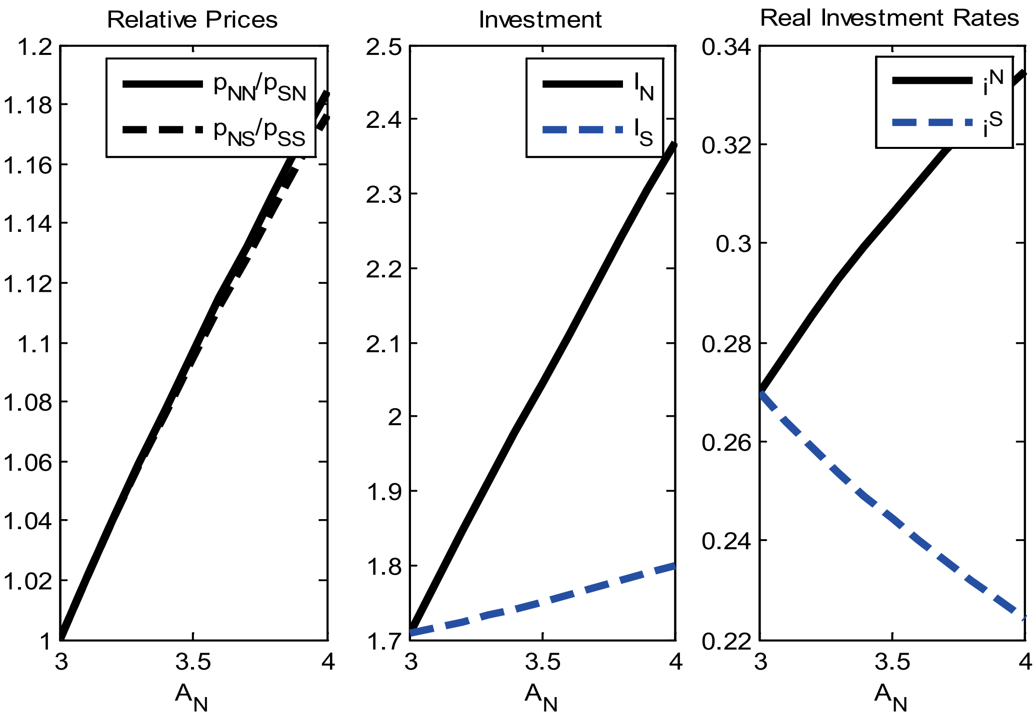
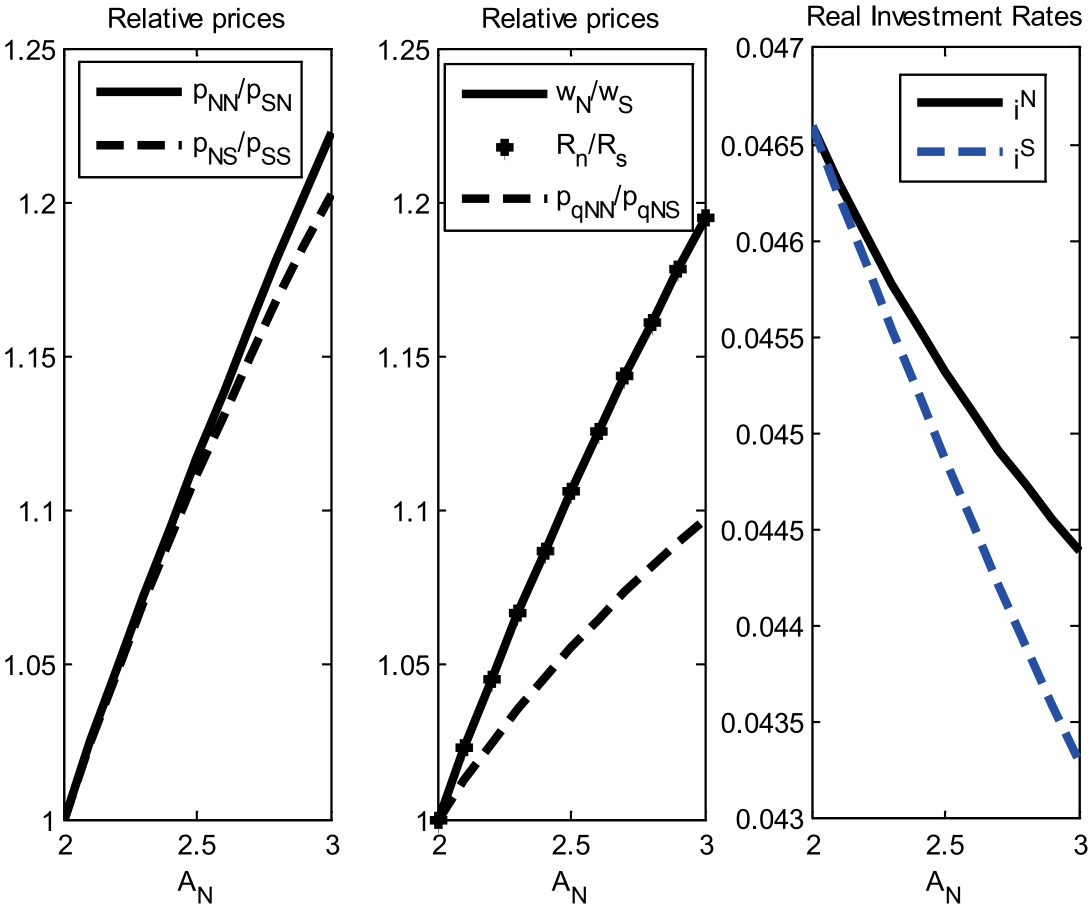


Figure I.D.1: Relative Prices and Purchases of Intermediate Investment Goods in Alternative Model with Investment Produced by a Quadratic Aggregator over Differentiated Intermediate Investment Goods.



References

Arkolakis, Costas. 2010. "Market Penetration Costs and the New Consumers Margin in International Trade." *Journal of Political Economy*. 118(6): 1151-1199.

Alessandria, George. and Joseph P. Kaboski. 2011. "Pricing-to-Market and the Failure of Absolute PPP," *American Economic Journal: Macroeconomics*, 3, 91–127.

Allen, Treb. 2012. "Information Frictions and Trade." Unpublished Manuscript.

Bergstrand, Jeffrey H. 1990. "The Heckscher-Ohlin-Samuelson Model, the Linder Hypothesis and the Determinants of Bilateral Intra-industry Trade." *Economic Journal* 100 (403): 1216–29.

Bhagwati, Jagdish N. 1984. "Why Are Services Cheaper in the Poor Countries?" *Economic Journal*, 94(374): 279–86.

Balassa, Bela. 1964. "The Purchasing-Power Parity Doctrine: A Reappraisal." *Journal of Political Economy*, 72(6): 584–96.

Barro, Robert J. 1991. "Economic Growth in a Cross Section of Countries." *Quarterly Journal of Economics*, 106(2): 407–43.

Burstein, Ariel, and Nir Jaimovich. 2008. "Understanding Movements in Aggregate and Product-level Real Exchange Rates." Unpublished.

Caselli, Francesco and James Feyrer. 2007. "The Marginal Product of Capital." *Quarterly Journal of Economics*, 122(2): 535-568.

Chaney, Thomas. 2008. "Distorted Gravity: The Intensive and Extensive Margins of International Trade." *American Economic Review*, 98(4): 1707-1721.

Crucini, Mario J., Chris I. Telmer, and Marios Zachariadis. 2005. "Understanding European Real Exchange Rates." *American Economic Review*, 95(3): 724–38.

Eaton, Jonathan, and Samuel Kortum**.** 2001. "Trade in Capital Goods." *European Economic Review*, 45(7): 1195–1235.

Engel, Charles. 1999. "Accounting for U.S. Real Exchange Rate Changes." *Journal of Political Economy*, 107(3): 507–38.

Fajgelbaum, Pablo, Gene M. Grossman, and Elhanan Helpman. 2011. "Income Distribution, Product Quality, and International Trade." *Journal of Political Economy*, 119(4): 721-765.

Fieler, Ana Cecilia. 2011. "Non-homotheticity and Bilateral Trade: Evidence and a Quantitative Explanation." *Econometrica* 79 (4): 1069–1101.

Fitzgerald, Doireann, and Stefanie Haller. 2012. "Pricing-to-Market: Evidence from Plant-Level Prices." Unpublished.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review* 98(1): 394-425.

Gopinath, Gita, Pierre-Oliver Gourinchas, Chang-Tai Hsieh, and Nicholas Li. 2011. "International Prices, Costs, and Markup Differences." *American Economic Review* 101: 2450-2486.

Gourio, Francois and Leena Rudanko. 2011. "Customer Capital." NBER Working Paper 17191.

Hallak, Juan Carlos. 2006. "Product Quality and the Direction of Trade." *Journal of International Economics*. 68(1): 238-265.

Harrod, Roy F. 1933. *International Economics*. London: James Nisbet & Co.

Heston, Robert. 2011. "Dwelling Services." International Comparison Program Book, Chapter 12.

Hsieh, Chang-Tai, and Peter J. Klenow. 2007. “Relative Prices and Relative Prosperity.” *American Economic Review*, 97(3): 562–85.

Hummels, David and Volodymyr Lugovsky. 2009. “International Pricing in a Generalized Model of Ideal Variety.” *Journal of Money, Credit, and Banking*, 41:3-33.

Hunter, Linda. 1991. “The Contribution of Non-homothetic Preferences to Trade.” *Journal of International Economics* 30, 345-358.

Khandelwal, Amit. 2010. “The Long and Short (of) Quality Ladders.” *Review of Economic Studies*. 77: 1450-1476

Krugman, Paul. 1987. “Pricing to Market When the Exchange Rate Changes.” in S. W. Arndt and J. Richardson, eds., Real Financial Linkages Among Open Economies, London: MIT Press.

Krugman, Paul. 1980. “Scale Economies, Product Differentiation, and the Pattern of Trade.” *American Economic Review* 70(5): 950-959.

Lancaster, Kelvin J .1966. "A New Approach to Consumer Theory." *Journal of Political Economy* 74(2): 132-157.

Manova, Kalina and Zhiwei Zhang. 2012. “Export Prices across Firms and Destinations.” *Quarterly Journal of Economics* 127:379-436.

Markusen, James R. 2010. “Putting Per-Capita back into Trade Theory and Policy.” NBER Working Paper 15903.

Matsuyama, Kiminori. 2000. “A Ricardian Model with a Continuum of Goods under Nonhomothetic Preferences: Demand Complementarities, Income Distribution, and North-South Trade.” *Journal of Political* Economy 108 (6): 1093–1120.

Melitz, Marc J. and Gianmarco I.P. Ottaviano. 2008. “Market Size, Trade, and Productivity.” *Review of Economic Studies* 75: 295-316.

Mitra, Devashish and Vitor Trindade. 2005. “Inequality and Trade.” *Canadian Journal of Economics* 38 (4): 1253–71.

Nakamura, Emi and Dawit Zerom. 2010. “Accounting for Incomplete Pass-Through.” *Review of Economic Studies*. 77: 1192-1230.

Ottavanio, Gianmarco. I. P., Takatoshi Tabuchi. and Jacques-François Thisse. 2002. “Agglomeration and Trade Revisited”, *International Economic Review*, 43, 409–436.

Samuelson, Paul A. 1964. “Theoretical Notes on Trade Problems.” *Review of Economics and Statistics*, 46(2): 145–54.

Simonovska, Ina. 2011. “Income Differences and Prices of Tradables.” UC Davis.

Summers, Robert, and Alan Heston. 1991. “The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988.” *Quarterly Journal of Economics*, 106(2): 327– 68.

Valentinyi, Ákos and Berthold Herrendorf. 2012. “Which Sectors Make Poor Countries So Unproductive?” *Journal of the European Economic Association*, 10(2): 323-341.

CHAPTER II

Does a Rising Tide Lift All Boats? Welfare Consequences of Asymmetric Growth

Abstract

A common presumption in macroeconomics and development economics is that increased growth in the aggregate enhances welfare for everyone in the economy. I show that instead, if the underlying growth is a productivity increase in the sector consumed primarily by one group, the welfare of a second group may fall. I demonstrate this effect in two cases. In the first case, skill-biased technological change in sectors consumed by the skilled rich increases their income beyond the increase in economic wealth, causing a decline in the consumption and welfare of the low-skilled poor. This result stands in contrast to the standard model of skill-biased technological change. The second case examines trade between two countries, and demonstrates circumstances under which an increase in productivity in the nontradable sector of one country causes a welfare decline for the other country. The paper discusses evidence in support of the effects in both cases. This analysis demonstrates that a rising tide need not lift all boats and that the precise nature of consumption patterns is important for welfare.

1. Introduction

Welfare is inherently difficult to measure, but evidence suggests that well-being of the poor has in fact fallen or stagnated in the face of economic growth. In the U.S., real GDP per capita increased 73% between 1970 and 2000, while the real wages of the lowest quintile earners decreased by over 20%.[1] Likewise, Brazilian GDP per capita increased over 46% during the same period, yet the living conditions of the poorest residents have not improved.[2] Thus the "rising tide" of economic growth did *not* by necessity "lift all boats", as JFK famously predicted in 1963.

Recent evidence also suggests that the "falling tide" of the recent economic crisis may have "lifted some boats". According to the World Bank Development Research Group, in 2008 (the year of the global financial crisis) the number and share of the population living on less than $1.25 a day fell in every part of the world for the first time on record. Preliminary estimates from 2010 suggest the decline has continued.[3] Similarly, data from the Gallup World Poll show that 132 million people became more food secure between between 2005 and 2008.[4] This evidence is seemingly difficult to reconcile with the fact that global output fell so substantially in 2008.

Indeed, an implicit assumption of standard economic models is that economic growth increases the welfare of everyone in the economy, even if growth is accompanied by increases in inequality. My analysis formally demonstrates circumstances under which growth actually lowers welfare for some. This paper examines two cases in which "rising tides" lead to "falling boats": First, sector-biased, skill-biased technological change in a closed economy, and second, service-sector productivity growth in a two-country model of international trade. In each case, a productivity increase complements labor supplied by a subset of people, and is biased toward the good consumed by that subset of people. As a result, the income of that subset rises more than the increase in aggregate economic wealth, causing a redistribution of productive resources away from goods consumed by others in the economy.

The analysis differs from earlier studies that have examined circumstances in which economic growth may reduce welfare. Examples include models of the Dutch Disease, as discussed in Corden and Neary (1982) and Krugman (1987), and of Immiserising Growth (see Baghwati 1958). Both Dutch Disease and Immiserising Growth rely on specific conditions that need not hold in general. In contrast, the explanation I provide in this paper focuses on

---

[1] More direct measures of welfare have also demonstrated stagnating well-being for America's poor during times of economic growth. For example, the United States Department of Agriculture documents that food security rates did not improve between 1995 (when they started reporting data) and the mid-2000s.

[2] According to the World Bank Development Research Group, in 2000, over 10% of the population continued to live on less than $1.25 a day.

[3] See *The Economist*, March 3rd-12th 2012, p. 81.

[4] See Headey (2012).

alternative conditions that apply to labor markets in a closed economy and to a two-country international trade setting. Furthermore, the models presented below directly address the “trickle down phenomenon” often heard in policy debates.

The first part of this paper examines the effects of skill-biased technological change in a closed economy. There is a growing consensus that new technologies complement skill, either directly or through productivity growth in the production of skill-complementing capital. Existing models of skill-biased technological change and capital-skill complementarity offer explanations for the rising skill premium in the latter half of the twentieth century and predict that wage inequality is likely to continue to increase. The analysis below expands these models to incorporate an additional insight, that new technologies appear to be directed not only toward factors of production (skilled labor and capital), but also toward goods consumed predominantly by the rich. The result of this asymmetric growth is a fall in the welfare of the low-skilled poor in addition to rising wage inequality. This finding is in contrast with the implications of the canonical one-sector model of skill-biased technological change, in which welfare increases despite rising wage inequality.[5]

An extensive literature has documented the failure of U.S. economic growth to “trickle down” to the lowest quintile of wage earners. Beaudry and Green (2003), for example, propose a model of organizational change that can generate falling real wages. However, their model relies on a counterfactual increase in the price of capital.[6] In contrast to the model in Beaudry and Green, the welfare implications of the model presented below do not rely on any assumptions about the existence or price of capital. Furthermore, the proposed model with sector-specific, skill-biased technological change is consistent with several features of the macroeconomy during the last half of the Twentieth Century, including 1) increasing expenditure shares of high-end services, 2) an increasing skill premium, 3) increasing skill intensity in high-end service sectors, and 4) a fall in the price of capital.

The proposed model of sector-specific, skill-biased technological change extends the one-sector, two-factor model in Acemoglu (1998) to an economy with two sectors producing

---

[5] See Acemoglu (1998, 2003)

[6] Similarly, in Caselli (1999), new machines that complement skilled workers replace old machines that complement unskilled workers. See Acemoglu (2002) for a survey.

Yachts and Potatoes and two types of agents (Rich and Poor).[7] Yachts represent goods or services consumed primarily by the Rich, while Potatoes represent goods and services consumed by the Poor. The Rich agents own an endowment of high-skilled labor, while the Poor own an endowment of low-skilled labor.[8] The key assumptions are, first, demand for Yachts is increasing in income; second, skill-biased technological improvements are sector specific; and third, the elasticity of substitution between high skilled labor and low skilled labor is greater than unity. If technology improves in the Yacht sector, the wage of the skilled Rich increases. The Rich in turn use their increased income to demand more Yachts, which requires skilled labor to flow out of the Potato sector and into the Yacht sector. The result is a fall in the supply of Potatoes. If preferences are strongly nonhomothetic such that the Poor consume only Potatoes, their welfare will decline.

The model feature that technological growth has been biased toward the goods that the Rich consume has some empirical support in the macroeconomic literature. Buera and Kaboski (2011) document that as income has grown in the latter half of the twentieth century, there has been a substantial increase in the expenditure share of skill-intensive services such as finance, insurance, real estate, and architectural services. This evidence suggests that demand for these services depends on wealth: as wealth increases, consumers shift toward consumption of skill-intensive services ("Yachts"). Furthermore, Jorgenson and Stiroh (2000) argue that the majority of TFP growth has been in the production of computers and IT, and Bosworth and Triplett (2000) show that the most intensive users of computer technology have included high-skill services such as finance, insurance, and communications.

This supporting empirical evidence is consistent with arguments in Acemoglu (1998, 2003) that technological change responds to market forces. As the rich demand more financial services, for example, the returns to the inputs in financial service production increase, which in turn increases the incentive to create software for the finance industry. The implication of this form of asymmetric growth, according to the model presented below, is a bifurcation of the

---

[7] Appendix C extends the capital-skill complementarity model in Krusell, Ohanian, Rios-Rull, and Violante (2000), which has three factors of production, and derives the same welfare implications. This paper considers capital-skill complementarity to be consistent with skill-biased technological change and therefore refers to the two interchangeably. In contrast to Beaudry and Green (2003), the results here are consistent with a fall in the price of capital over time.

[8] I use the terms 'high-skilled' and 'low-skilled' in accordance with the literature on skill-biased technological change. However, the model's mechanism is relevant when technological change is biased toward any factor (e.g. capital) that is not equally owned across groups.

economy: skilled labor flows from sectors consumed by the Poor to those consumed by the Rich, depriving the Poor of goods and services.

This pattern of bifurcation is equally salient among goods or services within the same sector. Broda and Romalis (2009) document that low-income households consume a basket of goods that is entirely different from the basket of high-income individuals, even though the goods are similarly classified. Their evidence is based on scanner data for consumer goods such as Maxwell House coffee and Starbucks, but a similar pattern is likely to hold for the service sector as well. For example, low-income households use basic medical services at local clinics while the wealthy undergo plastic surgeries. If we reinterpret Yachts to be high-end services such as cosmetic plastic surgeries, the model offers insights into the implications of a plastic surgeon's office obtaining state-of-the-art operating equipment: Skilled nurses leave the clinic in the poor neighborhood to earn a higher wage at the plastic surgeon's office in the wealthy neighborhood, driving up prices or reducing quality at the clinic.

This phenomenon also is consistent with the chronic underdevelopment of the poorest neighborhoods in America, South Africa, and elsewhere. If technological improvements have been biased toward investments in products for the wealthy, skilled labor and capital will flow into the provision of goods and services for the wealthy, leaving fewer productive inputs to provide for the poor. In the poorest neighborhoods, where goods and services are consumed exclusively by the low-skilled residents, only low-quality services provided by primarily low-skilled workers will remain. Imagine a state-of-the-art auto repair shop built near a gated community in Cape Town, South Africa. Skilled mechanics will earn a high return using the new equipment, leaving the low-skilled auto workers to repair cars for the poor out of shacks in the townships. Since the low-skilled mechanics work with inferior capital equipment their marginal product remains low, as does their income. Low income implies that demand for goods and services in low-income neighborhoods remains insufficient to attract new investments that would, in turn, increase wages and wealth.

Sector-specific, skill-biased technological change is not the only source of asymmetric growth that has implications for income, demand patterns, and welfare. The second part of the paper examines welfare effects of asymmetric growth in the context of international trade between two countries. Section 4 extends the basic framework from Balassa (1964) and Samuelson (1964) to permit imperfect substitutability between two countries' tradable goods.

The analysis demonstrates that productivity growth in the nontradable sector of one country may increase the price of tradables in another country, causing a fall the other country's welfare.

In the model below, country A and country B each produce a nontradable good and a tradable good that is an imperfect substitute for the other country's tradable. Productivity growth in A's nontradable sector causes a fall in welfare in B when the fall in the price of nontradables in A causes a shift in A's demand toward domestically produced tradables. This occurs whenever the elasticity of substitution between tradables and nontradables is greater than unity.[9] As the nontradable in A becomes cheaper due to increased productivity, the consumer in A substitutes toward nontradables and away from tradables. The increase in demand for A's nontradables lowers the value of tradables produced in B, causing a fall in B's terms of trade.

As in the model of sector-specific, skill-biased technological change, welfare falls in the two-country model due to a reallocation of factor inputs toward a sector (the service sector in country A) that is disproportionately consumed by one group (country A). This simple mechanism has interesting implications for prices and patterns of trade. For example, if China's demand shifts toward domestically produced goods as its productivity increases, U.S. imports from China will become more expensive, causing a fall in welfare in the U.S.

The two-country model may also help explain recent trends in global poverty. As mentioned at the beginning of the Introduction, in 2008 the number and share of the population living on less than \$1.25 a day fell in every part of the world for the first time on record. Similarly, 132 million people became more food secure between between 2005 and 2008. This evidence is seemingly difficult to reconcile with the fact that global output fell so substantially in 2008. However, it is fully consistent with the model's prediction that, as demand for resources falls in rich countries, certain goods may become more accessible to the poor. In other words, a "falling tide" may have "lifted some boats".

The two-country model builds on the framework in Balassa (1964) and Samuelson (1964) in that it allows for differences in productivity between a country's service sector and its tradable sector. Balassa and Samuelson observe that services tend to be more expensive in rich countries

---

[9] This assumption on the elasticity of substitution is a sufficient condition for the welfare effects mentioned above. Alternative assumptions may cause a decline in welfare, but I focus on the case of an above-unity elasticity of substitution because this is the implicit assumption in the new trade literature that focuses on differentiated goods under monopolistic competition.

than in poor countries. The Balassa-Samuelson hypothesis suggests that this correlation is because rich-country productivity is higher in the tradable sector than in the nontradable sector.

While Balassa-Samuelson is concerned with persistent differences in sectoral productivity, the model below examines the effects of a *change* in productivity in a country's service sector. When service-sector productivity increases in the rich country, the price of services in that country falls, consistent with Balassa-Samuelson. There is an additional effect in the model below, which is the increase in tradable prices in the poor country. This is due to the above-unity elasticity of substitution between tradables and nontradables and the imperfect substitutability between countries' tradable goods. The assumption of an above-unity elasticity of substitution is fully consistent with q-complementarity (as defined by Hicks 1970), and therefore is also consistent with notions that the marginal utility of services should increase with an increase in consumption of consumer goods, and vice-versa.

The models below do not incorporate endogenous technology. However, in reality technological improvements are likely to respond to market forces. For example, the arguments in Acemoglu (1998, 2003) imply that, in the closed-economy model of Section 3, technology should flow to the high-end sectors as demand for these products increases. If this endogenous technology response is skill-biased, then the welfare decrease for the Poor may be persistent and self-reinforcing in the absence of countervailing forces in the economy (such as Hicks-neutral growth and technology spillovers). Similarly, in the two-country model, as demand for services within a country increases (due, for example, to nonhomothetic preferences that place more weight on services as the economy grows), technology should flow to the service sector and prices of tradables should increase for other countries that do not experience service-sector technology improvements.

The remainder of the paper proceeds as follows: Sections 2 and 3 constitute the first part of the paper and are concerned with the closed-economy model of sector-specific, skill-biased technological change. Section 2 reports the evidence that technological improvements have been biased toward goods predominantly consumed by the wealthy. Section 3 details the baseline model and illustrates the welfare effects of sector-specific skill-biased technological change. Section 4 presents the two-country model of asymmetric technological change. Section 5 concludes.

2. Macroeconomic Evidence of Sector-Biased Technological Change

A near consensus has emerged that U.S. economic growth, especially in the 1990s, has primarily been due to productivity growth in the production and the use of information technology (IT) equipment.[10] To the extent that IT use is unevenly distributed across sectors, technological progress will be asymmetric. The questions I address in this section are first, whether there has been substantial asymmetry in the use of IT equipment (and therefore economic growth), and second, whether this asymmetry is related to consumption demand patterns.

Triplett and Bosworth (2000) note that IT use has, indeed, been concentrated in a handful of industries. The 1992 capital flow tables show that five industries (financial services, wholesale trade, business services, insurance, and communications) alone accounted for over half of new purchases of computers. If the measure of IT includes communications equipment in addition to computers and peripheral equipment, the air transportation industry also is included as a primary user of IT. The pattern based on the 1997 capital flow table is remarkably similar: At a more aggregated industry level, the three primary users of computers, software, and communications equipment are information, finance and insurance, and professional and technical services.

Of the IT-intensive industries mentioned above, four can be linked to NIPA consumption categories: finance, insurance, professional services, and air transportation. The expenditure share of each of these categories has increased in the latter part of the Twentieth Century; their combined share increased by over 57% between 1970 and 2000. As Buera and Kaboski (2011) document, each of these is a relatively skill-intensive service industry, and consumption categories that experienced increasing expenditure shares are almost exclusively skill-intensive services. Other categories, such as food, clothing, and low-skill services, have fallen or stagnated as a share of personal consumption expenditures. I interpret this evidence as indicative of nonhomothetic preferences: As income rises, demand shifts toward skill-intensive services, including those that are the most intensive users of IT.[11]

In this paper I therefore interpret evidence of productivity growth in the use and production of IT capital as technological change that is biased toward goods consumed by the Rich. Technological change is also assumed to be skill-biased based on the overwhelming

---

[10] See Stiroh (2002), Jorgensen (2001), Jorgensen and Stiroh (2000)

[11] The fact that the services demanded by the Rich are skill-intensive is irrelevant for the theoretical results presented below. The relevant fact is that new technology complements skilled labor.

evidence in support of skill-biased technological change (including capital-skill complementarity) in the latter part of the Twentieth Century.[12] Section 3 models skill-biased technological change in the simplest form by allowing IT technology to augment skill in production functions with two factors (skilled and unskilled labor). Appendix II.C treats IT equipment as an additional factor in production functions in which IT capital and skill are relative complements.

3. Baseline Model

The baseline model consists of two factors (high-skilled labor H and low-skilled labor L), two agent types (Rich and Poor), and two goods (Yachts and Potatoes) in a static economy. H Rich agents each inelastically supply one unit of high-skilled labor and L Poor agents each supply a unit of low-skilled labor. Here technology is taken as exogenous, and all markets are competitive.

3.1 Consumer Preferences

Rich (R) and Poor (P) consumers have identical nonhomothetic preferences over Yachts and Potatoes of the form

$$U_i(F_i, Y_i) = \max(a \times log(F_i + b), Y_i)$$

where $i \in R, P$ and Y_i is consumption of Yachts by consumer type i. I use F_i to denote consumption of Potatoes by consumer i (P already refers to Poor agents). This form of preferences has the useful property that consumption switches from exclusively Potatoes to exclusively Yachts as wealth crosses a certain threshold determined by the scale parameters a and b.[13] It captures the fact documented in Broda and Romalis (2009) that low-income households consume a basket of goods that is entirely different from the basket of high-income individuals, even though the goods may be similarly classified. For example, the Rich consume high-quality Starbucks coffee while the Poor consume Maxwell House instant coffee. The

---

[12] See, for example, Bound and Johnson (1992), Autor, Levy and Murnane (2003), and Autor, Katz, and Kearney (2008)

[13] A more common form of preferences in the structural change literature takes the form $U_i(F_i, Y_i) = \begin{cases} F_i \; if \; F_i < \overline{F} \\ \overline{F} + Y_i \; if \; F_i > \overline{F} \end{cases}$, in which the wealthy consume both Potatoes and Yachts but only once they've satiated their demand for Potatoes. The welfare implications are robust to this form of preferences, but these preferences are analytically inconvenient.

evidence in Broda and Romalis is based primarily on scanner data and applies mainly to different brands of goods within a sector, but I make the additional assumption that the Yacht bundle includes skill-intensive service sectors that are not included in the Potato bundle, such as financial planning services and architectural services. Since the skill-intensive services have experienced the majority of technological improvements in the form of IT use, I will assume that technological growth occurs primarily in the Yacht sector (see section 3.4).

An interesting quality of the consumer preferences is that if the wealth of the Poor were to increase, they would initially consume more Maxwell House coffee and Mickey's Malt Liquor (referred to collectively as Potatoes). At some point their wealth may be high enough that they instead purchase fine wines, airline tickets, and financial services (Yachts). I assume that endowments and technologies are such that the low-skilled Poor remain low-income and thus consume only Potatoes, while the high-skilled rich consume only Yachts. We can thus rewrite preferences as

$$U_R(\cdot) = Y_R \qquad U_P(\cdot) = a \times \log(F_P + b).$$

If the Rich were handed a Potato, it would not increase their utility. This seems reasonable; wealthy households likely have little use for malt liquor since it would take up cabinet space reserved for higher quality alcoholic beverages. Similarly, if the Poor were handed a Yacht their utility would not increase. This is clearly a less palatable assumption but may be appropriate in some contexts. If the low-income poor were given a claim on architectural services they could not use it without owning a home (which they may not be able to afford). Rather than actually use the service they would exchange it for a good or service that will provide them with utility.

3.2 Production

Potatoes (F) and Yachts (Y) are competitively produced with a constant-returns-to-scale technology using high-skilled labor and low-skilled labor:

$$F = F(z_F H_F, L_F)$$

and

$$Y = Y(z_Y H_Y, L_Y)$$

where H_j and L_j are high-skilled labor and low-skilled labor employed in sector $j \in F, Y$ and z_j is the skill-augmenting technology parameter in sector j. Here I assume that production has the same constant elasticity of substitution (CES) functional form as the models in Acemoglu (1998,2003):

$$F = \left[\eta(z_F H_F)^{\frac{\sigma_F - 1}{\sigma_F}} + (1-\eta)L_F^{\frac{\sigma_F - 1}{\sigma_F}}\right]^{\frac{\sigma_F}{\sigma_F - 1}}$$

and

$$Y = \left[\mu(z_Y H_Y)^{\frac{\sigma_Y - 1}{\sigma_Y}} + (1-\mu)L_Y^{\frac{\sigma_Y - 1}{\sigma_Y}}\right]^{\frac{\sigma_Y}{\sigma_Y - 1}}.$$

Appendix II.B examines equilibrium effects when production functional forms are not specified, and Appendix II.C incorporates IT capital into a nested CES functional form similar to that used in Krusell, Ohanian, Rios-Rull, and Violante (2000).

3.3 Equilibrium and the Effects of Asymmetric Growth

In the static competitive equilibrium consumers maximize utility subject to their budget constraints; firms maximize profits, and labor markets clear. The H Rich agents' collective budget constraint is

$$Hw_H \geq F_R p_R + Y_R p_Y$$

where w_H is the wage for high-skilled labor, p_j is the price of good j, and j_R is consumption of good j by Rich agents. Since endowments and technology are such that the Rich have enough wealth to exclusively purchase Yachts, their budget constraint can be written as

$$Hw_H \geq Yp_Y.$$

Furthermore, since production is competitive and exhibits constant returns to scale, p_Y will equal the cost-minimizing bundle of inputs necessary to produce one Yacht. Thus

$$Yp_Y = H_Y w_H + L_Y w_L$$

and we can rewrite a representative Rich agent's problem as

$$\max \left[\mu(z_Y H_Y)^{\frac{\sigma_Y - 1}{\sigma_Y}} + (1-\mu)L_Y^{\frac{\sigma_Y - 1}{\sigma_Y}}\right]^{\frac{\sigma_Y}{\sigma_Y - 1}}$$

$$\text{s.t.} \quad Hw_H \geq H_Y w_H + L_Y w_L . \qquad (1)$$

Likewise, the representative Poor agent's problem is

$$\max \quad .5 \times \log\left(\left[\eta(z_F H_F)^{\frac{\sigma_F-1}{\sigma_F}} + (1-\eta)L_F^{\frac{\sigma_F-1}{\sigma_F}}\right]^{\frac{\sigma_F}{\sigma_F-1}}\right)$$

$$\text{s.t.} \qquad Lw_L \geq H_F w_H + L_F w_L. \tag{2}$$

Viewing the consumers' problem as a choice over consumption of the two labor types is helpful for understanding the comparative static effects of an increase in z_Y, which in equilibrium will depend on the substitution elasticities σ_Y and σ_F. Equilibrium is fully characterized by the budget constraints (equations (1) and (2)), utility maximization by the Rich:

$$\frac{w_H}{w_L} = \frac{\mu}{1-\mu} z_Y^{\frac{\sigma_Y-1}{\sigma_Y}} \left(\frac{L_Y}{H_Y}\right)^{\frac{1}{\sigma_Y}}, \tag{3}$$

utility maximization by the Poor:

$$\frac{w_H}{w_L} = \frac{\eta}{1-\eta} z_F^{\frac{\sigma_F-1}{\sigma_F}} \left(\frac{L_F}{H_F}\right)^{\frac{1}{\sigma_F}}, \tag{4}$$

and market clearing:

$$L_F + L_Y = L, \tag{5}$$

$$H_F + H_Y = H. \tag{6}$$

As noted above, technological improvements have been biased toward high-end services, which are modeled here as Yachts. Therefore the object of interest is skill-biased technology in the Yacht sector, z_Y.

Proposition 1: If high-skill labor and low-skilled labor are substitutes in the production of Yachts ($\sigma_Y > 1$), then an increase in skill biased technology in the Yacht sector (z_Y) will cause a decrease in the amount of Potatoes produced and therefore a decline the welfare of the Poor. If labor types are substitutes in the production of Potatoes ($\sigma_F > 1$) the decline will be due to an outflow of high-skill labor from the Potato sector to the Yacht sector. If labor types are complements ($\sigma_F < 1$) in the production of Potatoes the decline will be due to an outflow of both inputs from the Potato sector.

Proof: See Appendix II.A.

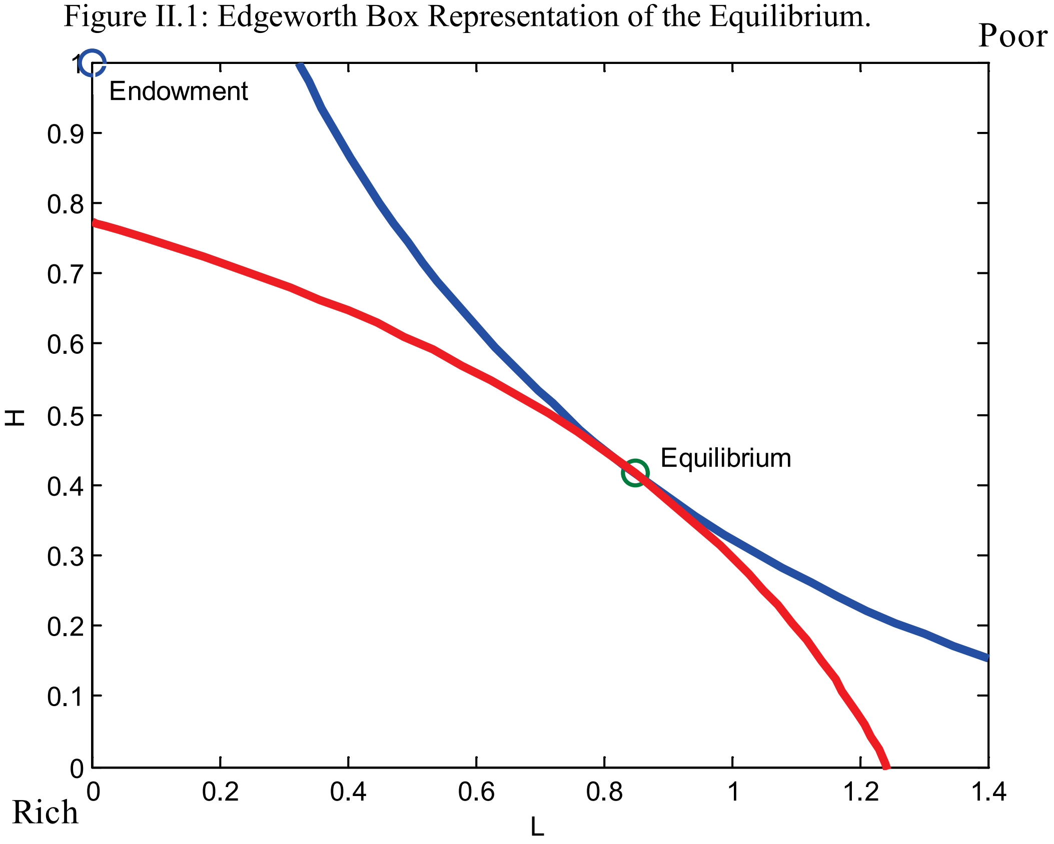
According to Katz and Murphy (1992), Angrist (1995), and Krusell, Ohanian, Rios-Rull, and Violante (2000), the empirically relevant case is when labor types are substitutes ($\sigma_Y > 1$

and $\sigma_F > 1$), although the crucial assumption for a fall in the welfare of the Poor is simply $\sigma_Y > 1$. An increase in z_Y drives up the wage premium, increasing the income of the Rich. Rich agents use their income to effectively purchase bundles of high-skill labor and low-skill labor. Since the $z_Y H_Y$ bundle is a substitute for L_Y in the Rich's utility function, the increase in z_Y increases $z_Y H_Y$, causing the Rich to desire a substitution of H_Y for L_Y. Since the increase in z_Y also increases the return to skilled labor and therefore the wealth of the Rich, the Rich are able to meet their desire for more skilled labor by purchasing skilled labor from the Poor. Skilled labor therefore flows from the Poor to the Rich (from the Potato sector to the Yacht sector).

The effect on the allocation of low-skilled labor, L, depends on the elasticity of substitution in the Potato sector. If $\sigma_F < 1$, labor types are complements for the Poor and the z_Y-induced decline in H_F lowers the value of L_F, which in turn diminishes the income of the Poor relative to the income of the Rich. In this case, the Rich have enough wealth to purchase more low-skilled labor in addition to high-skilled labor. If $\sigma_F > 1$, which is likely the empirically relevant case, the outflow of H from the Poor's consumption bundle causes a desire to substitute L for H, which increases the value of L relative to the case of complements. The Poor then are able to retain enough wealth to purchase low-skilled labor from the consumption bundle of the Rich.

When inputs are substitutes in each sector, the net effect is a fall in the utility of the Poor. This is because the effect of the outflow of high-skilled labor from the Potato sector outweighs the effect of the inflow of low-skilled labor (see Appendix II.A). Figures II.1 through II.3 illustrate the net effect of an increase in z_Y using an Edgeworth Box in which the representative agents trade high-skilled labor and low-skilled labor. Figure II.1 shows the initial equilibrium. The isoutility lines are identical to isoquants in the production of Potatoes for the Poor and Yachts for the Rich. Note that the original endowment of (L, H) to the Rich is $(0,1)$. The point labeled "Equilibrium" is the point of tangency between the isoutility lines of the Rich and the Poor, and the price vector is the line (not shown) between the "Endowment" point and the "Equilbrium" point.

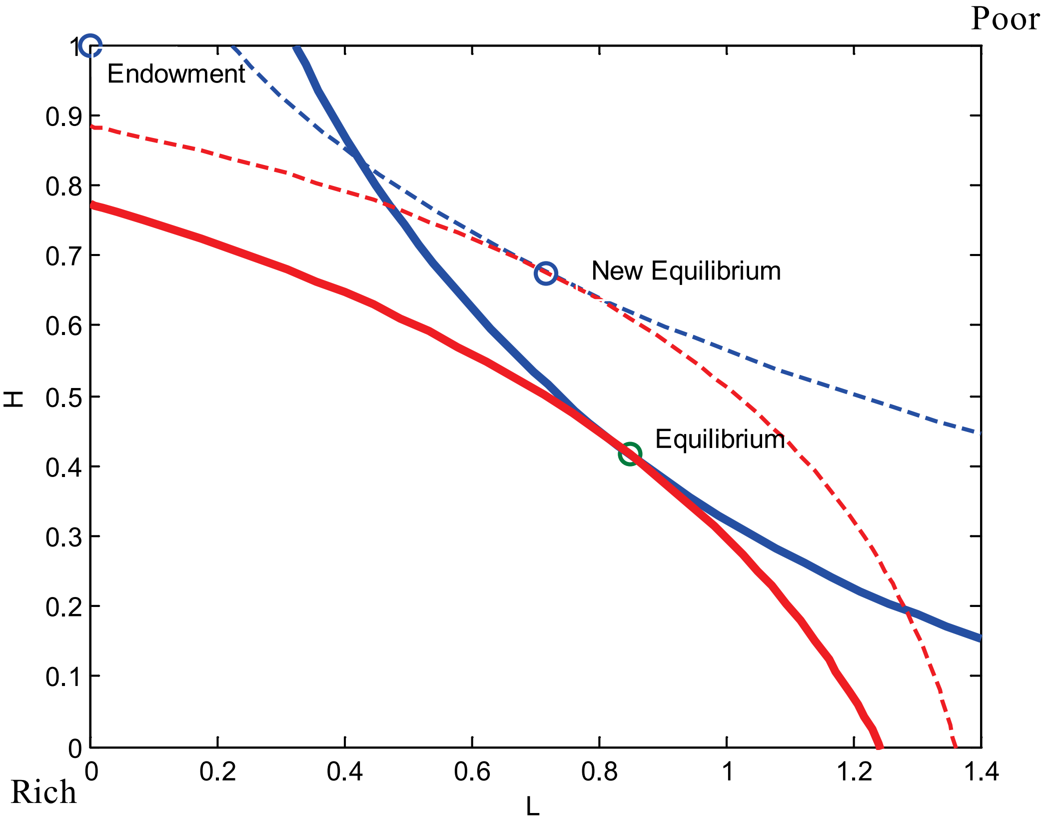
Figure II.1: Edgeworth Box Representation of the Equilibrium.



Note: The ratio of H to L is the same as the skill ratio in 2000 where high-skilled labor equals hours worked by college graduates.

Figure II.2 shows the effects of an increase in z_Y. The dashed lines are the new isoutility lines, and the point of tangency between the dashed lines is the new equilibrium. Note that the isoutility line of the Poor (the concave line) has shifted toward the point of zero consumption for the Poor, indicating that the utility of the Poor unambiguously falls when z_Y increases. The welfare decline of the Poor is due to the fact that z_Y complements the endowment of the Rich. The increased value of skilled labor, along with the substitution away from unskilled labor in the production of Yachts, allows the Rich to use their increased wealth to purchase additional skilled labor for the production of Yachts. Valuable skilled labor flows from Potato production to Yacht production, leaving the Poor worse off.

Figure II.2: Effect of an increase in Sector-Biased, Skill-Biased Technological Change.

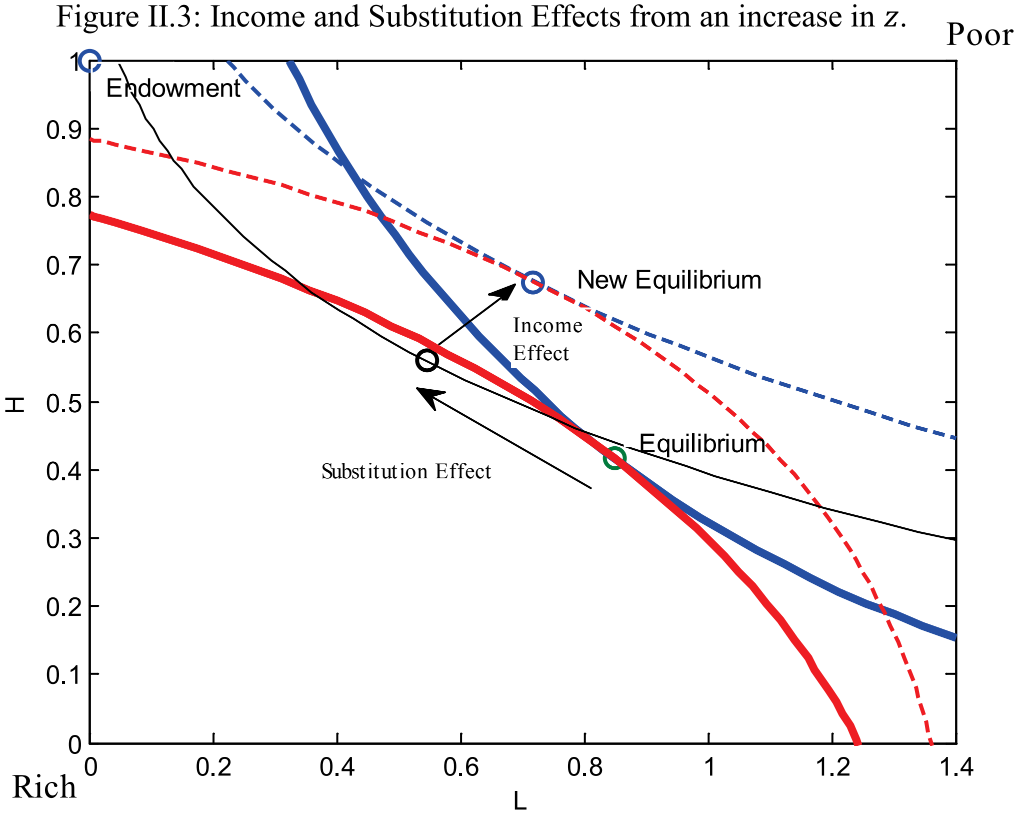


Note: The increase in z_Y is five-fold to illustrate the effects.

Figure 3 decomposes the change in allocations into what are labeled a *substitution effect* and an *income effect*. The substitution effect is defined as the change in allocations induced by an increase in z_Y when the economy's endowment point is assumed to be the original equilibrium (rather than (0,1)). The remaining distance from the original equilibrium to the actual new equilibrium is the income effect.

Figure 3 shows that the income effect, rather than the substitution effect, drives down the utility of the Poor. In fact, the substitution effect places both the Poor and the Rich on slightly higher isoutility lines (the thin solid lines). The income effect captures the fact the Rich are endowed with high-skilled labor, which has increased in value. The Rich are able to use their increased wealth to purchase additional skilled labor for the production of Yachts.

Figure II.3: Income and Substitution Effects from an increase in z.



An alternative way to understand the mechanism driving down the welfare of the Poor is through prices. If we normalize the price of low-skilled labor to unity, then the price of a Potato is

$$p_F = \left[\eta^{\sigma_F}\left(\frac{w_H}{z_F}\right)^{1-\sigma_F} + (1-\eta)^{\sigma_F}\right]^{\frac{1}{1-\sigma_F}}$$

and the price of a Yacht is

$$p_Y = \left[\mu^{\sigma_Y}\left(\frac{w_H}{z_Y}\right)^{1-\sigma_Y} + (1-\mu)^{\sigma_Y}\right]^{\frac{1}{1-\sigma_Y}}.$$

When z_Y increases the marginal product of skilled labor increases, driving up w_H. The price of Yachts falls because the increase in z_Y is greater than the increase in w_H ($dw_H/dz < 1$). Potatoes, meanwhile, do not benefit from price-reducing technological change, and thus the price of Potatoes increases because of the increase in w_H. Therefore the Poor do not benefit from higher wages but must pay a higher price for their consumption good.

The Poor would benefit if technological change augments either factor in the Potato sector or augments low-skilled labor in the Yacht sector. For example, technological change biased toward low-skilled labor in the Yacht sector pulls down w_H and the price of Potatoes relative to the return on low-skilled labor, thus improving the welfare of the Poor. The Poor likewise benefit from skill-biased technological change in the Potato sector: An increase in z_F increases w_H relative to w_L, but the overall effect is a fall in the price of Potatoes.

The greater is the elasticity of substitution in the Yacht sector, the greater is the consumption loss for the Poor in response to an increase in z_Y. Define $\hat{F} = dF/F$ and $\hat{z}_Y = dz_Y/z_Y$. Then total differentiation of equations 1 through 6 yields the response of Potato production to a small change in skill-biased technology in the Yacht sector:

$$\hat{F} = -F^{-\frac{\beta}{\beta-1}} \left[\frac{\sigma_Y H_Y L(\sigma_Y - 1)(1-\eta) L_F^{-\frac{1}{\sigma_F}} L_Y}{(\sigma_F H_F + \sigma_Y H_Y)\ L_F\ (\sigma_Y - 1) + (\sigma_F L_F + \sigma_Y L_Y)(H_F + \sigma_Y H_Y)} \right] \times \hat{z}_Y,$$

the magnitude of which is increasing in σ_Y when $\sigma_Y > 1$. Most estimates of the elasticity of substitution between skilled labor and low-skilled labor are between 1.4 and 2, implying that an increase in skill-biased technology in the Yacht sector drives down the supply of Potatoes.[14] Note that the direction of the change in the supply of Potatoes does not depend on factor intensities in the two sectors.[15]

3.4 Calibration

To get a sense of the magnitude of the consumption loss for the Poor I calibrate the model by choosing $\sigma_Y = \sigma_F = 1.4$, which is at the lower end of the empirical estimates of the elasticity of substitution between skilled and low-skilled labor. I set the skill ratio, $\frac{H}{L}$, equal to 0.7, which is close to the 2000 relative supply in Buera and Kaboski (2011) and the 1996 relative supply in Acemoglu (2002). The starting values for z_Y and z_F are equal to one. Finally, I choose $\eta = \mu = 0.62$ to match the wage premium in 2000, which is approximately 2.1 (see Acemoglu and Autor 2010). With these parameter values, a percent increase in z from a starting value of 1 causes a change in the supply of Potatoes of -0.1%.

Acemoglu (2002) postulates that skill biased technology increased almost tenfold in the U.S. between 1970 and 1990, based on a one-sector model with an elasticity of substitution equal

---

[14] See Katz and Murphy (1992), Angrist (1995), and Krusell, Ohanian, Rios-Rull, and Violante (2000).

[15] Also note that the model is not related to the Rybcszynski Theorem, which applies to changes in the amount of a factor that is available to both sectors.

to 1.4. The model here shows that if the full extent of technological improvements had been specific to sectors exclusively consumed by wealthy college graduates, the consumption loss for low-skilled workers would have been around a magnitude of 22% in the absence of hicks-neutral technological improvements, increases in the relative supply of high-skilled labor, technology spillovers, and other sources of economic growth.

The assumption that technological improvements are confined to the Yacht sector is illustrative but not realistic. If technological improvements occur in both sectors, the net effect on the supply of Potatoes will depend on the relative magnitude of skill-biased technological change in the Yacht sector. Table II.1 shows different combinations of increases in z_Y and z_F that achieve the same increase in the wage premium, along with the percent change in Potatoes per low-skilled worker. With an elasticity of substitution equal to 1.4 in both sectors, an 85% increase in z_Y requires a 13% increase in z_F to ensure that the supply of Potatoes does not fall. When the elasticity is 2, a 75% increase in z_Y will depress the Poor's consumption of Potatoes even if z_F increases by 22%. When growth is more symmetric, the consumption and welfare of the Poor improves.

Table II.1: Response of Consumption of the Poor to Skill Biased Technological Improvements

$\sigma_Y = \sigma_F = 1.4$			$\sigma_Y = \sigma_F = 2$		
Increase in z_Y	Increase in z_F	Change in F	Increase in z_Y	Increase in z_F	Change in F
100%	0%	-7.0%	100%	0%	-11.0%
95%	4%	-4.8%	95%	4%	-9.1%
90%	8%	-2.6%	90%	8%	-7.3%
85%	13%	0.0%	85%	12%	-5.4%
80%	18%	2.6%	80%	17%	-3.0%
75%	24%	5.7%	75%	22%	-0.7%
70%	29%	8.3%	70%	28%	2.1%
65%	36%	11.9%	65%	34%	4.9%
60%	43%	15.4%	60%	41%	8.2%

Note: Each row generates an equivalent increase in the skill premium for a given value of the elasticity of substitution.

As in the canonical one-sector model in Acemoglu (1998), inequality increases in all cases. For example, when z_Y doubles the skill premium increases by 23.5%. Furthermore, the expenditure share of Yachts increases by over 15%. The model therefore matches both the increasing trend in the skill premium and the trend documented in Buera and Kaboski (2009) of an increasing expenditure share of high-end services over time. Finally, the model predicts that the skill intensity of the Yacht sector should increase in response to the increase in z_Y due to the inflow of skilled labor. This is exactly the pattern observed in the high-end sectors mentioned in Section 2. Between 1940 and 2000, the average skill intensity over all sectors increased almost 70%, while the average skill intensity in the high-end service sectors increased over 250%.[16]

High-end services are not the only goods consumed predominantly by the Rich. For example, the Rich may consume different auto repair and personal care services than do the Poor based on the geographic location of these services relative to where the Rich live. The model above predicts that if an auto repair shop in a Rich neighborhood experiences an increase in skill-intensive technology (or capital, as demonstrated in the Appendix), then the availability of repair services in Poor neighborhoods will fall Such a phenomenon may help explain the chronic underdevelopment of neighborhoods in South Africa during periods of aggregate economic growth, and the prevalence of retail deserts in poor American neighborhoods (see Schuetz, Kolko, and Meltzer 2012).

For a concrete example of how asymmetric growth contributes to retail deserts, consider the recent experience of North Park Hill, a poor neighborhood that is almost entirely lacking in grocery stores. In 2009, during the depths of the Great Recession, the healthy-food grocer Sunflower Markets planned to open a store in North Park Hill.[17] In 2010, during the economic recovery, Sunflower withdrew its plans for a Park Hill store and began looking for locations in wealthier communities.[18] While Sunflower executives have not explicitly stated that increasing wealth in other parts of the city (and hence higher demand) were responsible for retracting their plans for a Park Hill store, their decision to do so is consistent with the predictions of the model above. It is also consistent with the experiences of other poor American neighborhoods, many of which rely on tax transfers in the form of grants to community development organizations to finance local grocery stores (see the bottom of the article cited in Footnote 16).

---

[16] Based on IPUMS data used in Buera and Kaboski (2011). The high-end industries include Security and commodity brokerage and investment companies, Banking and credit, Legal services, Engineering and architectural services, Real estate, Insurance, and Air transportation.

[17] The Denver Post, March 9, 2009 Front Page http://www.denverpost.com/business/ci_11867922?source=bb

[18] The Denver Post, March 10, 2010 http://www.denverpost.com/ci_14644222

4. Rising Tides in International Trade

This section examines the effects of biased technological change in the form of a productivity increase in a country's service sector. As in Section 3, growth is asymmetric in the sense that the productivity increase (1) complements labor supplied by a subset of people, and (2) is exclusive to a sector consumed by that subset of people. In the case of international trade, the subset of people who benefit from the asymmetric growth is the residents of the country with an increase in its service-sector productivity.

The model below features two countries, A and B, each with a representative consumer who has utility over a nontraded service and a bundle of traded consumer goods. Each country produces a unique traded good that is an imperfect substitute for the traded good of the other country. There are no barriers to trade, and all goods are produced under perfect competition. The production technology for all goods is linear in labor, and labor is inelastically supplied in both countries.

4.1 Consumer Preferences

The utility of the representative consumer from country $j \in \{A, B\}$ is

$$U_j = \left(S_j^{\frac{\sigma-1}{\sigma}} + C_j^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}, \tag{7}$$

Where S_j is the agent j's consumption of the service produced in country j. The aggregate good C_j consists of tradable goods from A and B:

$$C_j = \left[X_j^{\frac{\gamma-1}{\gamma}} + Y_j^{\frac{\gamma-1}{\gamma}}\right]^{\frac{\gamma}{\gamma-1}},$$

where X_j is consumption in j of the tradable good produced in country A and Y_j is consumption in j of the tradable good produced in country B. The elasticity of substitution between the service and the aggregate consumer good is σ, and the elasticity of substitution between the tradable goods is γ.

The budget constraint of agent j is

$$w_j L_j = p_{Sj} S_j + p_X X_j + p_Y Y_j, \tag{8}$$

where p_{Sj} is the price of S_j, p_X is the price of the tradable good X produced in A and p_B is the price of the tradable good Y produced in country B. Market clearing for the tradable goods implies $X = X_A + X_B$ and $Y = Y_A + Y_B$.

Utility maximization subject to (8) yields optimal consumption ratios for a given set of prices:

$$\frac{Y_j}{X_j} = \left(\frac{p_X}{p_Y}\right)^{\gamma}, \qquad \frac{C_j}{S_j} = \left(\frac{p_{Sj}}{p_C}\right)^{\sigma}, \tag{8}$$

where p_C is the cost-minimizing price of the bundle C of the tradable goods:

$$p_C = \left(p_X^{1-\gamma} + p_Y^{1-\gamma}\right)^{\frac{1}{1-\gamma}}.$$

Note that both countries face the same prices p_X and p_Y for tradable goods, so the price p_C of the aggregate consumer good is also equal in both countries.

4.2 Production

The technologies in j for production of services and tradables are

$$S_A = zL_{SA}, \qquad S_B = L_{SB}, \qquad X = L_{XA}, \qquad Y = L_{YB},$$

where L_{SA} and L_{XA} are A's labor allocations to its service and tradable sectors, and L_{SB} and L_{YB} are B's labor allocations to its service and tradable sectors. Without loss of generality, labor supply is assumed to be equal to L in both countries. S_B, X, and Y are produced at unit labor cost, while S_A is produced at labor cost $1/z$. Country A's service sector is permitted to have non-unity labor productivity because the comparative static of interest is a change in z.

All sectors are perfectly competitive, so prices equal marginal costs. The wage in B is the numeraire, so that prices can be written:

$$p_{SA} = \frac{w}{z}, \qquad p_X = w, \qquad p_{SB} = 1, \qquad p_Y = 1. \tag{9}$$

where w is the wage in country A.

4.3 Equilibrium

Equilibrium consists of optimal consumption ratios in each country (8), labor market clearing in each country,

$$L = \frac{S_A}{z} + X_A + X_B, \qquad L = S_B + Y_A + Y_B, \tag{10}$$

and the trade balance condition

$$wX_B = Y_A, \tag{11}$$

where equation (11) follows from substituting the labor market clearing conditions into each country's budget constraint.

4.3 Welfare Effects of Asymmetric Growth

Proposition 2: Welfare in B is falling in z when tradable goods are imperfect substitutes and the elasticity of substitution between the nontradable and the tradable bundle, σ, is greater than unity.

Proof: Appendix II.D.

If the elasticity of substitution between tradables is less than unity, it is still possible for welfare to fall in B. I focus on the case of substitutability because it is consistent with the implicit assumption in the macro and trade literatures that focus on the effects of monopolistic competition.

Intuitively, the result in Proposition 2 is a result of a fall in B's terms of trade when A's service-sector productivity increases. An increase in z lowers the price of nontradable services in A. Since the elasticity of substitution between services and tradables is greater than unity, country A responds to the price decrease by demanding more services and fewer tradables. This lowers the value of tradables relative to the value of labor in A, causing A to devote labor resources to the production of services. Therefore, resources are reallocated away from provision of tradables for trade with B to the provision of nontradable services for A.

An alternative way to conceptualize the result in Proposition 2 is through price effects. The consumer price index in B can be written

$$P_B = \left(1 + (w^{1-\gamma} + 1)^{\frac{1-\sigma}{1-\gamma}}\right)^{\frac{1}{1-\sigma}}, \tag{11}$$

which is increasing in w. Therefore, if the value of labor in A increases relative to the value of labor in B, the consumer price index in B will increase and B's welfare will fall. Under the conditions stipulated in Proposition 2, the increase in z causes an increase in w because country A indirectly demands more of its labor relative to the labor of country B.

4.4 Discussion of the Two-Country Model.

The analysis in Section 4 suggests that productivity growth in any country's service sector may increase prices in other countries, reducing their welfare. This result is fairly general in that the conditions leading to this result are imperfect substitutability between traded goods and an above-unity elasticity of substitution between nontraded services and tradables. Indeed, these assumptions are implicit in most macro and trade models that incorporate monopolistic competition.

The model presented above offers a solution to what appears to be a puzzling decline in abject poverty in all parts of the world following the global economic crisis in 2008: A fall in productivity in sectors (such as financial sectors) that do not serve the world's poorest members may have freed up economic resources to provide goods and services to the poor by lowering the price of these goods and services.

One implication is that an increase in global economic activity may have adverse consequences for the poorest people in the world. An additional implication of the model is that as service-sector productivity increases in developing countries, developed countries will experience higher prices and lower welfare in the absence of other forms of economic growth.

5. Conclusion

This paper has presented two cases in which productivity gains not only fail to "trickle down" to everyone in the economy, but actually lower the welfare of a group of people. In each case, the paper presents evidence in support of the assumptions underlying the models, as well as anecdotal evidence that lend support to the models' welfare implications.

The key mechanism driving the result in the models above is an increase in the income of a subset of people that supersedes the increase in aggregate wealth in response to economic growth. This mechanism does not operate in standard models that explain rising inequality. In the canonical model of skill-biased technological change, for example, economic gains are disproportionately directed to a subset of people but nonetheless improve everyone's welfare.

The primary difference between the models presented above and models of inequality is that the models presented above assume that consumption bundles not identical across the population.

In the case of sector-biased, skill-biased technological change, the difference in consumption bundles between the skilled Rich and the unskilled Poor is due to nonhomothetic preferences. The welfare decline of the low-skilled Poor is greater the more biased is growth toward high-end goods and services, or toward services in wealthy neighborhoods. The welfare loss is also greater the less important are high-end goods and services in the consumption bundle of the Poor. In the extreme, the Poor do not consume any goods that experience productivity gains, and their consumption/welfare loss is substantial.

In the case of international trade, consumption bundles differ across countries due to nontraded services, as in Balassa (1964) and Samuelson (1964). Under the assumptions that tradable goods are imperfect substitutes, and that the elasticity of substitution between services and tradables is greater than unity, the country without the service-sector productivity increase experiences a decline in its terms of trade and in its welfare.

The focus of this paper is on illustrating mechanisms that are likely to apply in a number of contexts rather than on developing methods to test these models empirically. A rigorous empirical test of the models would require matching consumption of disaggregated goods and services to the inputs used in production of the specific goods and services. Existing data clearly are inadequate for such an analysis. The development of suitable datasets for testing this proposition is beyond the scope of this paper.

An interesting avenue for empirical research would be to document productivity gains and the adoption of new technology by service establishments at the neighborhood level. If service establishments in high-income neighborhoods experience skill-biased technology improvements or utilize more high-tech capital than their counterparts in low-income areas, then the model in this paper may help explain the chronic underdevelopment of some of the poorest neighborhoods.

Appendix II.A

Proof of Proposition 1: Total differentiation of equations (1) through (6) yields

$$\widehat{H_F} = -\frac{(\sigma_Y - 1)(\sigma_F L_F + L_Y)H_Y}{H_F[(\sigma_F L_F + \sigma_Y L_Y)] + H_Y[((\sigma_F + \sigma_Y - 1)L_F + \sigma_Y L_Y)]}\widehat{z_Y} \tag{A1}$$

and

$$\widehat{L_F} = \frac{(\sigma_F - 1)(\sigma_Y - 1)\sigma_Y H_Y L_Y}{(\sigma_F H_F + \sigma_Y H_Y)\ L_F\,(\sigma_Y - 1) + (\sigma_F L_F + \sigma_Y L_Y)(H_F + \sigma_Y H_Y)}\widehat{z_Y}, \tag{A2}$$

where $\hat{x} = dx/x$ for any variable x. Equation (A1) implies that dH_F/dz is negative if and only if $\sigma_Y > 1$. Assuming this is the case, (A2) implies that dL_F/dz is negative if and only if $\sigma_F < 1$. If we assume that the elasticity of substitution is greater than unity in both sectors, then an increase in z will cause an outflow of skilled labor from the Potato sector and an inflow of unskilled labor. We can determine the net effect on the supply of Potatoes by total differentiation of the Potato production function:

$$\frac{dF}{F}\,F^{\frac{\sigma_F}{\sigma_F-1}} = \eta K_F^{-\frac{1}{\sigma_F}} dK_F + (1-\eta)L_F^{-\frac{1}{\sigma_F}} dL_F.$$

Substituting in (A1) and (A2) yields

$$\hat{F} = -F^{-\frac{\sigma_F}{\sigma_F-1}}\left[\frac{\sigma_Y H_Y L(\sigma_Y - 1)(1-\eta)L_F^{-\frac{1}{\sigma_F}} L_Y}{(\sigma_F H_F + \sigma_Y H_Y)\ L_F\,(\sigma_Y - 1) + (\sigma_F L_F + \sigma_Y L_Y)(H_F + \sigma_Y H_Y)}\right] \times \hat{z},$$

which states that if $\sigma_Y > 1$ the supply of Potatoes decreases whenever there is an improvement in skill-biased technological change in the Yacht sector.

Appendix II.B

Here I alter the model in Section 3 to allow production of Yachts and Potatoes to use a general constant-returns-to-scale functional form. In the static competitive equilibrium consumers maximize utility subject to their budget constraints; firms maximize profits, and labor markets clear. The H Rich agents solve

$$\max \ Y_R$$

$$\text{s.t. } Hw_H = Yp_Y \tag{B1}$$

Likewise, the Poor agents solve

$$\max 1.5 \times \log(F_P + 1)$$

$$\text{s.t.} \quad Lw_L = Fp_F \tag{B2}$$

Prices of Potatoes and Yachts are equal to unit costs c_F and c_Y:

$$c_F(w_H, w_L) = p_F \tag{B3}$$

$$c_Y\left(\frac{w_H}{z}, w_L\right) = p_Y. \tag{B4}$$

Market clearing implies

$$H_F + H_Y = H \tag{B5}$$

$$L_F + L_Y = L. \tag{B6}$$

Shepard's Lemma determines conditional factor demands in the Food sector:

$$\frac{\partial c_F}{\partial w_H} = \frac{H_F}{F} \tag{B7}$$

$$\frac{\partial c_F}{\partial w_L} = \frac{L_F}{F} \tag{B8}$$

and relative factor demands in the Yacht sector are derived setting marginal rates of technical substitution equal to the ratio of input prices:

$$\frac{\partial Y}{\partial H} / \frac{\partial Y}{\partial L} = \frac{r}{w} \tag{B9}$$

Equations A1 through A9 characterize the equilibrium. We can log-linearize the equilibrium equations to determine the effects of an increase in z on all endogenous variables:

$$\widehat{w_H} = \hat{Y} + \widehat{p_Y} \tag{B11}$$

$$\widehat{w_L} = \hat{F} + \widehat{p_F} \tag{B12}$$

$$\phi_H \widehat{w_H} + \phi_L \widehat{w_L} = \widehat{p_F} \tag{B13}$$

$$\theta_H \widehat{w_H} + \theta_L \widehat{w_L} = \theta_Z \hat{z} + \widehat{p_Y} \tag{B14}$$

$$\lambda_{H_F} \widehat{H_F} + \lambda_{H_Y} \widehat{H_Y} = 0 \tag{B15}$$

$$\lambda_{L_F} \widehat{L_F} + \lambda_{L_Y} \widehat{L_Y} = 0 \tag{B16}$$

$$\widehat{H_F} = \hat{F} + \phi_H \sigma_F (\widehat{w_L} - \widehat{w_H}) \tag{B17}$$

$$\widehat{L_F} = \hat{F} + \phi_L \sigma_F (\widehat{w_H} - \widehat{w_L}) \tag{B18}$$

$$\widehat{L_Y} - \widehat{H_Y} + (\sigma_Y - 1)\hat{z} = \sigma_Y (\widehat{w_H} - \widehat{w_L}) \tag{B19}$$

For any variable x above, $\hat{x} = \frac{dx}{x}$. I denote cost shares of labor in the Potato sector as $\phi_H \equiv w_H H_F / F p_F$ and $\phi_L \equiv w_L L_F / F p_F$. Likewise in the Yacht sector $\theta_H \equiv w_H H_Y / Y p_Y$ and $\theta_L \equiv w_L L_Y / Y p_Y$. The shares of labor types in each sector are $\lambda_{H_F} = \frac{H_F}{H}$, $\lambda_{H_Y} = \frac{H_Y}{H}$, $\lambda_{L_F} = \frac{L_F}{L}$, and $\lambda_{L_Y} = \frac{L_Y}{L}$. The elasticity of substitution between the labor types is σ_F in the Potato sector. In the Yacht sector σ_Y in the elasticity of substitution between zH and L. Solving the above system of equations yields the percentage change in Potatoes in response to a percentage increase in Yacht-specific, skill-biased technological change:

$$\hat{F} = -\frac{(\sigma_Y - 1)\lambda_{H_Y}}{\left\{(1 + \sigma_F)\lambda_{H_F} + \lambda_{H_Y}\left[\frac{\sigma_Y}{\phi_H} + \frac{\lambda_{L_F}}{\lambda_{L_Y}}\left(\frac{\sigma_F \phi_L}{\phi_H} - 1\right)\right]\right\}}\hat{z}$$

Potato production will fall in response to an increase in z whenever $\sigma_Y > 1$ and $\frac{\sigma_Y}{\phi_H} + \frac{\lambda_{L_F}}{\lambda_{L_Y}}\left(\frac{\sigma_F \phi_L}{\phi_H} - 1\right) > 0$. This latter condition will hold when production functions are of the CES form as in Section 3.

Appendix II.C

Here I extend the model in Section 3 to include equipment capital, K. Production of Potatoes and Yachts takes the nested CES form:

$$F = \left[\eta\left(\lambda\, K_F^{\frac{\sigma_F-1}{\sigma_F}} + (1-\lambda)H_F^{\frac{\sigma_F-1}{\sigma_F}}\right)^{\frac{\sigma_F}{\sigma_F-1}\frac{\beta-1}{\beta}} + (1-\eta)L_F^{\frac{\beta-1}{\beta}}\right]^{\frac{\beta}{\beta-1}}$$

$$Y = \left[\mu\left(\lambda\,(zK_Y)^{\frac{\sigma_Y-1}{\sigma_Y}} + (1-\lambda)H_Y^{\frac{\sigma_Y-1}{\sigma_Y}}\right)^{\frac{\sigma_Y}{\sigma_Y-1}\frac{\gamma-1}{\gamma}} + (1-\mu)L_Y^{\frac{\gamma-1}{\gamma}}\right]^{\frac{\gamma}{\gamma-1}},$$

which is similar to the production function estimated by Krusell, Ohanian, Rios-Rull, and Violante (2000). The technology parameter, z, augments capital in the Yacht sector only. Alternatively, we could assume that capital is sector-specific, and that productivity improvements are unique to the production of capital used in the yacht sector. With competitive markets the effects on factor demands and prices will be the same; the only difference is that capital in the Yacht sector would be measured as zK_Yinstead of K_Y. Krusell et al. implicitly assume that the unit of measurement of capital is zK (theirs is a one-sector model) in order to account for the fall in the price of equipment capital during the latter part of the Twentieth Century. However, this assumption is not necessary: in the calibrated general equilibrium model the price of K (and the price of zK_Y) can fall in response to an increase in z, as we demonstrate below.

Preferences are the same as in the baseline model, and we assume that the Rich own the economy's endowment of capital K in addition to high-skilled labor H. The representative Rich agent therefore solves

$$\max\ Y$$

$$\text{s.t.}\quad rK + w_H H \geq rK_Y + w_H H_Y + w_L L_Y \tag{C1}$$

and the Poor agent solves

$$\max\ .5 \times \log(F)$$

$$\text{s.t.}\quad w_L L \geq rK_F + w_H H_F + w_L L_F, \tag{C2}$$

where r is the price of capital.

In the competitive equilibrium the marginal rates of technical substitution must equal input prices. This consists of two equations in the Yacht sector,

$$\frac{\mu}{1-\mu}\lambda\, z^{\frac{\sigma_Y-1}{\sigma_Y}}\left(\lambda\,(zK_Y)^{\frac{\sigma_Y-1}{\sigma_Y}}+(1-\lambda)H_Y^{\frac{\sigma_Y-1}{\sigma_Y}}\right)^{\frac{\gamma-\sigma}{\gamma(\sigma-1)}} L_Y^{\frac{1}{\gamma}}K_Y^{-\frac{1}{\sigma_Y}}=\frac{r}{w_L} \tag{C3}$$

and

$$\frac{\mu}{1-\mu}\,(1-\lambda)z^{\frac{\gamma-1}{\gamma}}\left(\lambda\,(zK_Y)^{\frac{\sigma_Y-1}{\sigma_Y}}+(1-\lambda)H_Y^{\frac{\sigma_Y-1}{\sigma_Y}}\right)^{\frac{\gamma-\sigma}{\gamma(\sigma-1)}} L_Y^{\frac{1}{\gamma}}H_Y^{-\frac{1}{\sigma_Y}}=\frac{w_H}{w_L}, \tag{C4}$$

and two equations in the Food sector,

$$\frac{\eta}{1-\eta}\lambda\left(\lambda\,K_F^{\frac{\sigma_F-1}{\sigma_F}}+(1-\lambda)H_F^{\frac{\sigma_F-1}{\sigma_F}}\right)^{\frac{\beta-\sigma}{\beta(\sigma-1)}} L_F^{\frac{1}{\beta}}\,K_F^{-\frac{1}{\sigma_F}}=\frac{r}{w_L} \tag{C5}$$

and

$$\frac{\eta}{1-\eta}(1-\lambda)\left(\lambda\,K_F^{\frac{\sigma_F-1}{\sigma_F}}+(1-\lambda)H_F^{\frac{\sigma_F-1}{\sigma_F}}\right)^{\frac{\beta-\sigma}{\beta(\sigma-1)}} L_F^{\frac{1}{\beta}}\,H_F^{-\frac{1}{\sigma_F}}=\frac{w_H}{w_L}. \tag{C6}$$

Equations (C1) through (C6), in addition to market clearing conditions

$$K=K_F+K_Y,\qquad H=H_F+H_Y,\quad L=L_F+L_Y, \tag{C7-C9}$$

fully characterize the competitive equilibrium.

I calibrate the model using the parameter estimates in Krusell et al (2000):

$$\beta=\gamma=1.67,\qquad \sigma_Y=\sigma_F=0.67,\qquad \lambda=0.553,\quad \eta=0.587.$$

As in section 3, I set $\frac{H}{L}=.7$ to match its value in 2000. I set $\mu=0.65$ instead of 0.587 to help match the 2000 skill premium, $\frac{w_H}{w_L}=2.1$, and because a higher value of μ increases the relative skill intensity in the Yacht sector, consistent with the evidence in Buera and Kaboski (2009). The capital stock, $K=7$, is chosen to match the skill premium. The starting value for z is 1.

Table II.C1 shows the response of endogenous variables to a 10% increase in z. The supply of Potatoes, F, falls by 0.43%, due entirely to an outflow of skilled labor. Unskilled labor and capital actually flow into the Potato sector. When z increases, the technology-capital bundle zK_Y increases in the Yacht sector. Since zK_Y and H_Y are relative complements (determined by the magnitude of σ_Y relative to γ), the Rich demand more skilled labor in the Yacht sector, which increases w_H and H_Y. The Rich also demand less capital because the level of zK_Y is high relative to H_Y, which lowers the price of capital. The result is an outflow of capital from the Yacht sector and into the Potato sector. The stronger the relative complementarity between capital and skill, the stronger is the fall in r and in inflow of capital to the Potato sector. If baseline calibration is changed slightly to decrease the relative complementarity (through either an increase in σ_Y or a decrease in γ), the sign of the change in r ,K_F , or both can reverse. All other variable changes are robust to a wide range of parameter values.

Table II.C1: Response of Endogenous Variables to a 10% Increase in z

$\widehat{w_H}$	$\hat{r}$	$\widehat{K_F}$	$\widehat{H_F}$	$\widehat{L_F}$	$\hat{F}$	$\hat{Y}$
1.59%	-1.70%	0.51%	-1.45%	0.29%	-0.43%	2.07%

Note: The price of low-skilled labor, w_L, is normalized to 1.

Appendix II.D

This section derives the result in Proposition 2 from Section 5. A sufficient statistic for welfare in B is given by the consumer price index (11). Therefore the comparative static of interest is dw/dz. If $dw/dz > 0$, then the productivity increase causes an increase in B's consumption basket and a fall in its welfare. To determine dw/dz, I first simplify some of the equilibrium conditions.

Substitute the prices (9) into the tradable demand ratios (8), to yield

$$Y_j = X_j w^{\gamma}, \tag{D1}$$

$$C_j = X_j (1 + w^{\gamma-1})^{\frac{\gamma}{\gamma-1}} \tag{D2}$$

for $j = \{A, B\}$. The optimal demands of services in A and B from (8) are

$$S_A = C_A\,(w^{1-\gamma} + 1)^{\frac{\sigma}{1-\gamma}} \left(\frac{w}{z}\right)^{-\sigma}, \tag{D3}$$

$$S_B = C_B\,(w^{1-\gamma} + 1)^{\frac{\sigma}{1-\gamma}}. \tag{D4}$$

Solve the labor market clearing conditions (10) for S_A and S_B and substitute into (D3) and (D4). Also substitute (D2) for C_A and C_B:

$$z(L - X_A - X_B) = X_A (1 + w^{\gamma-1})^{\frac{\gamma}{\gamma-1}} (w^{1-\gamma} + 1)^{\frac{\sigma}{1-\gamma}} \left(\frac{w}{z}\right)^{-\sigma}, \tag{D5}$$

$$L - Y_A - Y_B = X_B (1 + w^{\gamma-1})^{\frac{\gamma}{\gamma-1}}\,(w^{1-\gamma} + 1)^{\frac{\sigma}{1-\gamma}}. \tag{D6}$$

Substitute the demand ratios (D1) for Y_B:

$$z(L - X_A - X_B) = X_A (1 + w^{\gamma-1})^{\frac{\gamma}{\gamma-1}} (w^{1-\gamma} + 1)^{\frac{\sigma}{1-\gamma}} \left(\frac{w}{z}\right)^{-\sigma}, \tag{D7}$$

$$L - Y_A - X_B w^{\gamma} = X_B (1 + w^{\gamma-1})^{\frac{\gamma}{\gamma-1}}\,(w^{1-\gamma} + 1)^{\frac{\sigma}{1-\gamma}}. \tag{D8}$$

Finally, substitute out X_B using the trade balance condition (11), and substitute out Y_A using (D1):

$$z\big(L - X_A(1 - w^{\gamma-1})\big) = X_A (1 + w^{\gamma-1})^{\frac{\gamma}{\gamma-1}} (w^{1-\gamma} + 1)^{\frac{\gamma}{1-\gamma}} \left(\frac{w}{z}\right)^{-\sigma}, \tag{D9}$$

$$L - X_A w^{\gamma} (1 - w^{\gamma-1}) = X_A w^{\gamma-1} (1 + w^{\gamma-1})^{\frac{\gamma}{\gamma-1}}\,(w^{1-\gamma} + 1)^{\frac{\gamma}{1-\gamma}}. \tag{D10}$$

Equations (D9) and (D10) fully characterize the equilibrium. There are two equations and two unknowns (w and X_A). To obtain dw/dz, totally differentiate (D9) and (D10), solve for dX_A, and substitute the second totally differentiated equation into the first. The total derivative is evaluated around the equilibrium for which the two countries are identical ($z = 1$). The result is

$$\frac{dw}{dz} = \frac{\sigma - 1}{2^{\frac{\sigma-1}{\gamma-1}}\gamma + \frac{3\sigma + \gamma}{2} - 1},$$

which is strictly positive whenever $\sigma > 1$ and $\gamma < \infty$.

References

Acemoglu, D. (1998) "Why do New Technologies Complement Skills? Directed Technical Change and Wage Inequality." *Quarterly Journal of Economics* 111:779-804.

Acemoglu, D. (2003) "Patterns of Skill Premia," *Review of Economic Studies* 69:781-809.

Acemoglu, D. 2002. "Technical Change, Inequality, and the Labor Market." *Journal of Economic Literature*, 40(1): 7-72.

Acemoglu, D. and D. Autor (2010) "Skills, Tasks and Technologies: Implications for Employment and Earnings." NBER Working Paper 16082.

Angrist, J. (1995) "The Economic Returns to Schooling in the West Bank and Gaza Strip." *American Economic Review*, 85(5): 1065-1087.

Autor, D., L. Katz, and M. Kearney (2008) "Trends in U.S. Wage Inequality: Revising the Revisionists." *Review of Economic Statistics* 90(2): 300-323.

Autor, D., F. Levy, and R. Murnane (2003) "The Skill Content of Recent Technological Change: An Empirical Investigation." *Quarterly Journal of Economics* 118: 1279–1333.

Balassa, Bela. 1964. "The Purchasing-Power Parity Doctrine: A Reappraisal." *Journal of Political Economy*, 72(6): 584–96.

Bhagwati, J. (1958) "Immiserizing Growth: A Geometrical Note," *Review of Economic Studies* 25: 201-205.

Beaudry, P. and D. Green (2003) "Wages and Employment in the United States and Germany: What Explains the Differences?" *American Economic Review,* 93(3): 573–602.

Bound, J. and G. Johnson (1992) "Changes in the Structure of Wages in the 1980s: An Evaluation of Alternative Explanations." *American Economic Review* 82(3) 371-92.

Broda, C. and J. Romalis (2009) "Welfare Implications of Rising Price Dispersion" University of Chicago mimeo.

Buera, J. and J. Kaboski (2011) "The Rise of the Service Economy." Forthcoming *American Economic Review.*

Caselli, F (1999) "Technological Revolutions." *American Economic Review*, 89(1): 78-102.

Corden, WM and JP Neary (1982). "Booming Sector and De-industrialisation in a Small Open Economy". *The Economic Journal* 92 (December): 825–848.

Headey, Derek (2012) "The Impact of the Global Food Crisis on Self-Assessed Food Security." International Food Policy Research Institute.

Hicks, John (1970) "Elasticity of Substitution Again: Substitutes and Complements." *Oxford Economic Papers*, 22(3): 289-296.

Jorgenson, D. (2001) "Information Technology and the U.S. Economy." *American Economic Review*, 91(1): 1-32.

Jorgenson, D. and K. Stiroh (2000) "Raising the Speed Limit: U.S. Economic Growth in the Information Age," *Brookings Papers on Economic Activity* 1: 125-236.

Katz, L. and K. Murphy (1992) "Changes in Relative Wages: Supply and Demand Factors." *Quarterly Journal of Economics* 107:35-78.

Krugman, P. (1987) "The Narrow Moving Band, the Dutch Disease, and the Competitive Consequences of Mrs. Thatcher." *Journal of Development Economics* 27: 41-55.

Krusell, P., Ohanian, L., Rios-Rull, J, and G. Violante (2000) "Capital-Skill Complemantarity and Inequality: A Macroeconomic Analysis." *Econometrica* 68, 1029-1053.

Samuelson, Paul A. 1964. "Theoretical Notes on Trade Problems." *Review of Economics and Statistics*, 46(2): 145–54.

Schuetz, Jenny, Jed Kolko, and Rachel Meltzer (2012), "Are Poor Neighborhoods 'Retail Deserts'", *Regional Science and Urban Economics*. 42: 269-285.

Stiroh, K. (2002) "Information Technology and the Productivity Revival: What Do the Industry Data Say?" *American Economic Review*, 92(5): 1559-1576.

Triplett, J. and B. Bosworth (2000) "Productivity in the Services Sector." The Brookings Institution. Mimeo.

CHAPTER III

The Role of Inventories and Speculative Trading in the Global Market for Crude Oil

Abstract

We develop a structural model of the global market for crude oil that for the first time explicitly allows for shocks to the speculative demand for oil as well as shocks to flow demand and flow supply. The speculative component of the real price of oil is identified with the help of data on oil inventories. Our estimates rule out explanations of the 2003-08 oil price surge based on unexpectedly diminishing oil supplies and based on speculative trading. Instead, this surge was caused by unexpected increases in world oil consumption driven by the global business cycle. There is evidence, however, that speculative demand shifts played an important role during earlier oil price shock episodes including 1979, 1986, and 1990. Our analysis implies that additional regulation of oil markets would not have prevented the 2003-08 oil price surge. We also show that, even after accounting for the role of inventories in smoothing oil consumption, our estimate of the short-run price elasticity of oil demand is much higher than traditional estimates from dynamic models that do not account for price endogeneity.

1. Introduction

There is no consensus in the academic literature on how to model the global market for crude oil. One strand of the literature suggests that the price of oil is determined by desired stocks. In this interpretation, shifts in the expectations of forward-looking traders are reflected in changes in the real price of oil and changes in oil inventories. Another strand of the literature views the price of oil as being determined by shocks to the amount of oil coming out of the ground ("flow supply of oil") and the amount of oil being consumed ("flow demand for oil") with little attention to the role of inventories. Much of the early literature on oil supply shocks is in that tradition, as are more recent economic models linking the real price of oil to fluctuations in the global business cycle.[1] Recently, there has been increasing recognition that both stock demand and flow demand for oil matter in modeling the real price of oil (see, e.g., Dvir and Rogoff 2010; Hamilton 2009a,b; Kilian 2009a; Alquist and Kilian 2010). In this paper, we propose a structural vector autoregressive (VAR) model of the global market for crude oil that explicitly nests these two explanations of the determination of the real price of oil and allows us to quantify the effects of different oil demand and supply shocks. Drawing on insights from the economic theory for storable commodities, we design a set of identifying assumptions that allows us to estimate jointly the expectations-driven component of the real price of oil and the components driven by flow demand and flow supply.

Constructing such an econometric model is nontrivial because the potential presence of a forward-looking component in the real price of oil considerably complicates the identification of the structural shocks. Traditional oil market VAR models implicitly equate market expectations with the econometric expectations formed on the basis of past data on oil production, global real activity, and the real price of oil. If traders respond to information about future demand and supply conditions not contained in the past data available to the econometrician, however, market expectations will differ from the expectations constructed by the econometrician, rendering traditional models of flow supply and flow demand invalid. For example, the desirability of holding oil stocks may change in response to news about oil discoveries or as traders anticipate wars in the Middle East or global recessions, or as traders respond to increased uncertainty about future oil supply shortfalls. None of these expectations shifts can be captured using standard models of flow demand and flow supply. In sections 2 and 3, we show that this problem can be overcome with the help of data on above-ground crude oil inventories. The intuition is that – unless the price elasticity of oil demand is zero – any expectation of a shortfall of future oil supply relative to future oil demand not already captured by flow demand and flow supply shocks necessarily causes an increase in the demand for above-ground oil inventories and

---

[1] See, e.g., Baumeister and Peersman (2012); Hamilton (2009a,b); Kilian (2008a, 2009a,b); Kilian and Murphy (2012).

hence in the real price of oil (see, e.g., Alquist and Kilian 2010; Hamilton 2009a). We refer to such a shock as a speculative demand shock in the spot market for crude oil. It is this type of shock that many researchers and policymakers explicitly or implicitly appeal to when attributing higher spot prices to speculation. We stress that such speculative demand shocks cannot be inferred directly from observables and can only be identified within the context of a structural econometric model.

Our definition of speculation in this paper is general in that we treat anyone buying crude oil not for current consumption, but for future use as a speculator from an economic point of view. Speculative purchases of oil usually occur because the buyer is anticipating rising oil prices. This anticipation may arise because of changes in expected fundamentals, for example, or because the buyer is anticipating other market participants' actions. Speculative purchases may also be precautionary in that they reflect increased uncertainty about future demand or supply conditions (see Alquist and Kilian 2010).

We do not take a stand on whether such speculative behavior is desirable from a social point of view. In particular, we do not attempt to distinguish between normal and excessive levels of speculation nor do we define speculation on the basis of who is trading or what positions these traders take. As discussed in Fattouh et al. (2012), there is no operational definition of excessive speculation in the literature. Rather we will show in section 4 that all speculative trading in the spot market for crude oil combined had little effect on the real price of oil between 2003 and mid-2008, making the distinction between normal and excessive levels of speculation moot at least for this episode.

Our analysis allows new insights into the genesis of historical oil price fluctuations. It is of particular relevance for recent policy discussions about the potential role of speculation in oil markets after 2003. First, as discussed in section 4, our estimates rule out speculation as a cause of the surge in the real price of oil between 2003 and mid-2008. Furthermore, under the maintained assumption of arbitrage between spot and futures markets, the absence of speculative pressures in the spot market implies that there cannot have been speculative pressures in the oil futures market either. This implication is fully consistent with other, more direct evidence on the impact of financial investors on oil futures prices (see Fattouh et al. 2012). Instead our model implies that both spot and futures prices during 2003-08 were driven by unexpected increases in world oil consumption. From this result we infer that additional regulation of oil futures markets would not have prevented the increase in the real price of oil in the spot market, which is the ultimate concern of policymakers.[2]

Second, although speculative trading does not explain the recent surge in the real price of oil, we show that it played an important role in several earlier oil price shock episodes. For example, it was a central feature of the oil price surge of 1979, following the Iranian Revolution, consistent with the narrative evidence in Barsky and Kilian (2002), and it helps explain the sharp decline in the real price of oil in early 1986 after the collapse of OPEC. It also played a central role in 1990, following Iraq's invasion of Kuwait. Although neither negative flow supply shocks nor positive speculative demand shocks alone can explain the oil price spike and oil inventory behavior of 1990/91, their combined effects do.

Third, our analysis sheds new light on the evolution of the real price of oil since 1978. We document that unexpected fluctuations in global real activity explain nearly the entire surge in the real price of oil between 2003 and mid-2008, even acknowledging that negative flow supply shocks raised the real price of oil slightly. Business cycle factors were also responsible for the bulk of the 1979/80 oil price increase in conjunction with sharply rising speculative demand in the second half of 1979. In contrast, flow supply shocks played only a minor role in 1979. The continued rise in the real price of oil in 1980 reflected negative flow supply shocks (caused in part by the outbreak of the Iran-Iraq War) as much as continued (if slowing) global growth, amidst declining speculative demand. Finally, there is evidence that the recovery of the real price of oil starting in 1999, following an all-time low in post-war history, was aided by coordinated supply cuts. Although our analysis assigns more importance to oil supply shocks than some previous studies, we conclude that, with the exception of 1990, the major oil price shocks were driven primarily by oil demand shocks.

Much of the prima facie case against an important role for speculation in oil markets rests on the fact that there has been no noticeable increase in the rate of inventory accumulation in recent years. Our model suggests that even after controlling for the effect of other shocks on crude oil inventories there is no evidence of rising speculative demand after 2003. Recently, Hamilton (2009a) pointed out that, as a matter of theory, speculative trading in oil futures markets may cause a surge in the real price of oil without any change in oil inventory holdings if

---

[2] It is important to note that our conclusion regarding the determinants of the spot price of oil would remain equally valid, if there were limits to arbitrage between spot and futures markets. Limits to arbitrage of course would undermine the argument that financial speculation is driving the spot price.

the short-run price elasticity of demand for gasoline is literally zero. Thus, it is essential that we pin down the value of this elasticity. We provide a theoretical model that shows that under reasonable assumptions about the oil refining industry, the short-run price elasticity of gasoline demand is about as high as the short-run price elasticity of oil demand. Hamilton (2009a) observed that existing estimates of the latter elasticity are so close to zero that the possibility of an elasticity of zero deserves further examination. These existing elasticity estimates, however, are based on dynamic reduced-form regressions that ignore the endogeneity of the real price of oil. They have no structural interpretation and suffer from downward bias. In section 5, we address this concern with the help of our structural VAR model. Not only does this model allow the construction of a direct estimate of the elasticity parameter based on exogenous shifts of the oil supply curve along the oil demand curve, but it also incorporates for the first time changes in oil inventories in computing the price elasticity of oil demand. Our posterior median estimate of the short-run price elasticity of oil demand of -0.26 is four times higher than standard estimates in the literature and there is little probability mass close to zero.[3] Thus, the limiting case of a zero price elasticity of demand discussed by Hamilton (2009a) is an unlikely explanation of 2003-08 surge in the real price of oil. The concluding remarks are in section 6.

2. VAR Methodology

Our analysis is based on a dynamic simultaneous equation model in the form of a structural VAR. Let y_t be a vector of endogenous variables including the percent change in global crude oil production, a measure of global real activity, the real price of crude oil, and the change in global crude oil inventories above the ground.[4] All data are monthly. The sample period is 1973.2-2009.8. We remove seasonal variation by including seasonal dummies in the VAR model.

2.1. Data

Our measure of fluctuations in global real activity is the dry cargo shipping rate index developed in Kilian (2009a). This real activity index is a business cycle index and stationary by construction. It is designed to capture shifts in the global use of industrial commodities. For more

---

[3] Even higher oil demand elasticity estimates have been obtained independently by other recent studies employing structural estimation methods. None of these studies account for changes in inventories, however.

[4] Unlike above-ground oil inventories that can be drawn down at short notice, oil below the ground (also known as reserves) is inaccessible in the short run and not available for consumption smoothing. Thus, it must be differentiated from oil inventories in the usual sense. We do not utilize data for reserves because no reliable time series data exist on the quantity of oil below the ground and because reserves data are not required for our identification. We discuss, however, how speculation based on below-ground inventories would be recorded within our model framework and how it may be detected in section 4.3.

details on the rationale, construction and interpretation of this index the reader is referred to the related literature. While there are other indices of global real activity available, none of these alternative proxies is as appropriate for our purpose of capturing shifts in the global demand for industrial commodities. Data on global crude oil production are available in the *Monthly Energy Review* of the Energy Information Administration (EIA). These data also include lease condensates, but exclude natural gas plant liquids. Oil production is expressed in percent changes in the model. The real price of oil is defined as the U.S. refiners' acquisition cost for imported crude oil, as reported by the EIA, extrapolated from 1974.1 back to 1973.1 as in Barsky and Kilian (2002) and deflated by the U.S. consumer price index. We use the refiners' acquisition cost for imported crude oil because that price is likely to be a better proxy for the price of oil in global markets than the U.S. price of domestic crude oil which was regulated during the 1970s and early 1980s. Following Kilian (2009a), the real price of oil is expressed in log-levels.[5]

Given the lack of data on crude oil inventories for other countries, we follow Hamilton (2009a) in using the data for total U.S. crude oil inventories provided by the EIA. These data are scaled by the ratio of OECD petroleum stocks over U.S. petroleum stocks, also obtained from the EIA. That scale factor ranges from about 2.23 to 2.59 in our sample.[6] We express the resulting proxy for global crude oil inventories in changes rather than percent changes. One reason is that the percent change in inventories does not appear to be covariance stationary, whereas the change in inventories does. The other reason is that the proper computation of the oil demand elasticity, as discussed below, requires an explicit expression for the change in global crude oil inventories in barrels. This computation is only possible if oil inventories are specified in changes rather than percent changes. Preliminary tests provided no evidence of cointegration between oil production and oil inventories.

---

[5] It is not clear a priori whether the real price of oil should be modeled in log-levels or log-differences. The level specification adopted in this paper has the advantage that the impulse response estimates are not only asymptotically valid under the maintained assumption of a stationary real price of oil, but robust to departures from that assumption, whereas incorrectly differencing the real price of oil would cause these estimates to be inconsistent. The potential cost of not imposing unit roots in estimation is a loss of asymptotic efficiency, which would be reflected in wider error bands. Since the impulse response estimates presented below are reasonably precisely estimated, this is not a concern in this study. It should be noted, however, that historical decompositions for the real price of oil rely on the assumption of covariance stationarity and would not be valid in the presence of unit roots.

[6] Petroleum stocks as measured by the EIA include crude oil (including strategic reserves) as well as unfinished oils, natural gas plant liquids, and refined products. The EIA does not provide petroleum inventory data for non-OECD economies. We treat the OECD data as a proxy for global petroleum inventories. Consistent series for OECD petroleum stocks are not available prior to 1987.12. We therefore extrapolate the percent change in OECD inventories backwards at the rate of growth of U.S. petroleum inventories. For the period 1987.12-2009.8, the U.S. and OECD petroleum inventory growth rates are reasonably close with a correlation of about 80%.

2.2. A Model of the Global Market for Crude Oil

The reduced-form model allows for two years worth of lags. This approach is consistent with evidence in Hamilton and Herrera (2004) and Kilian (2009a) on the importance of allowing for long lags in the transmission of oil price shocks and in modeling business cycles in commodity markets. The corresponding structural model of the global oil market may be written as

$$B_0 y_t = \sum_{i=1}^{24} B_i y_{t-i} + \varepsilon_t, \tag{1}$$

where ε_t is the vector of orthogonal structural innovations and $B_i, i = 0,...,24,$ denotes the coefficient matrices. The seasonal dummies have been suppressed for notational convenience. The vector ε_t consists of four structural shocks: The first shock corresponds to the classical notion of an oil supply shock as discussed in the literature ("flow supply shock"). This shock incorporates supply disruptions associated with exogenous political events in oil-producing countries as well as unexpected politically motivated supply decisions by OPEC members as well as other flow supply shocks. Second, we include a shock to the demand for crude oil and other industrial commodities that is associated with unexpected fluctuations in the global business cycle ("flow demand shock"). The third shock captures shifts in the demand for above-ground oil inventories arising from forward-looking behavior not otherwise captured by the model ("speculative demand shock").[7] Finally, we include a residual shock designed to capture idiosyncratic oil demand shocks not otherwise accounted for (such as weather shocks, changes in inventory technology or preferences, or politically motivated releases of the U.S. Strategic Petroleum Reserve).

Each of these shocks has unique characteristics. For example, an unexpected disruption of the flow of oil production (embodied in a shift to the left of the contemporaneous oil supply curve along the oil demand curve, conditional on all past data) raises the real price of crude oil and lowers global real activity within the same month. The impact effect on oil inventories is ambiguous. On the one hand, a negative flow supply shock will cause oil inventories to be drawn down in an effort to smooth consumption. On the other hand, the same shock may raise demand for inventories to the extent that a negative flow supply shock triggers a predictable increase in

---

[7] An alternative and less common view is that speculation may also be conducted by oil producers who choose to leave oil below the ground in anticipation of rising prices. The latter form of "speculative supply shock" would be associated with a negative flow supply shock in our framework rather than the building of above-ground inventories. Both forms of speculation are permitted by our model, but only speculation by oil traders is explicitly identified.

the real price of oil. Which effect dominates is unclear ex ante, so we do not restrict the sign of the impact response of inventories.

In contrast, a positive flow demand shock (embodied in a shift to the right of the contemporaneous oil demand curve along the oil supply curve, conditional on all past data) raises the real price of oil and stimulates global oil production within the same month. As in the case of a negative flow supply shock, the impact effect on inventories is ambiguous ex ante. Flow demand and flow supply shocks in turn differ from speculative demand shocks.

Given that crude oil is storable, the real price of oil also depends on the demand for oil inventories. This means that we must allow the price of oil to jump in response to any news about *future* oil supply or *future* oil demand that is not already embodied in flow supply and flow demand shocks. For example, upward revisions to expected future demand for crude oil (or downward revisions to expected future production of crude oil), all else equal, will increase the demand for crude oil inventories in the current period, resulting in an instantaneous shift of the contemporaneous demand curve for oil along the oil supply curve and an increase in the real price of oil. Such shifts could arise, for example, because of the anticipation of political unrest in oil-producing countries in the Middle East, because of the anticipation of peak oil effects, or because of the anticipated depletion of oil reserves. Likewise, traders may anticipate a global recession in the wake of a financial crisis, may anticipate higher future oil production as new deep sea oil is discovered off the shores of Brazil, or may anticipate the resumption of oil production in Iraq, as the stability of that country improves.

Rather than being associated only with future oil supply conditions or only with future oil demand conditions, speculative demand shocks reflect the expected shortfall of future oil supply relative to future oil demand. A positive speculative demand shock will shift the demand for above-ground oil inventories, causing in equilibrium the level of inventories and the real price of oil to increase on impact. The accumulation of inventories is accomplished by a reduction in oil consumption (reflected in lower global real activity) and an increase in oil production, as the real price increases.[8] Both flow demand shocks and speculative demand shocks have an expectational

---

[8] Although oil producers could conceivably react to the same news that triggers a positive speculative demand shock by lowering oil production in anticipation of predictable increases of the real price, there is no evidence that oil producers have responded systematically in this way. Instead, anecdotal evidence suggests that oil producers such as Saudi Arabia often have increased their production levels following positive speculative demand shocks, consistent with the view that the expected impact response should be weakly positive. Our analysis is based on the premise that these shocks are mutually uncorrelated, but allows multiple shocks to occur at the same time.

component. The feature that distinguishes flow demand shocks from speculative demand shocks is that positive flow demand shocks necessarily involve an increase in the demand for consumption in the current period, whereas speculative demand shocks do not.[9]

News about the level of future oil supplies and the level of future demand for crude oil are but one example of shocks to expectations in the global market for crude oil. An unexpected increase in the uncertainty about future oil supply shortfalls would have much the same effect. This point has been demonstrated formally in a general equilibrium model by Alquist and Kilian (2010). The main difference is that uncertainty shocks would not be associated with expected changes in future oil production or real activity. Finally, speculative demand shocks may also arise because traders' perception of what other traders think evolves or simply because of changes in beliefs not related to expected fundamentals. One of the attractive features of the econometric model is that it does not require the user to specify how expectations are formed.

2.3. Why Do We Not Include the Oil Futures Spread?

The focus in this paper is on modeling the real price of oil in the spot market. We do not explicitly model the oil futures market. Indeed, conceptually the futures market is distinct from the spot market. As discussed in Alquist and Kilian (2010) and Hamilton (2009a,b), there is an arbitrage condition linking the oil futures market and the spot market for crude oil. To the extent that speculation drives up the price in the oil futures market, arbitrage will ensure that oil traders buy inventories in the spot market in response. Thus we can focus on quantifying speculation in the spot market with the help of the oil inventory data without loss of generality.[10] In fact, the analysis in Alquist and Kilian implies that data on the oil futures spread are redundant in our structural VAR model, given that we have already included changes in above-ground oil inventories. The fact that the inclusion of oil inventory data makes oil futures prices redundant is particularly advantageous considering that oil futures markets were created only in the 1980s, and thus oil futures prices do not exist for a large part of our sample. Equally importantly, our model remains well specified even if the arbitrage between spot and futures markets were less than perfect at times, whereas a model including the oil futures spread would become invalid in

---

[9] For a theoretical analysis of flow demand shocks and how their effect on the real price of oil may be amplified by index funds trading in oil futures markets also see Sockin and Xiong (2012).

[10] This result breaks down only if demand for oil is completely price inelastic, a case that we discuss in more detail in section 5.

that case. Finally, inventory data allow us to impose identifying information about the price elasticity of oil demand, which could not be imposed when using the oil futures spread.

2.4. How Accurate are the Oil Inventory Data?

While our structural VAR specification has many advantages, it relies on the global crude oil inventory data being accurate. Two concerns regarding the reliability of these data stand out. First, much has been made of media reports that some speculators in 2007/08 leased oil tankers to store oil. Although the EIA does not provide data on the use of tankers for storage, leasing tankers is expensive and the extent to which tankers have been used for storage appears small and limited to the very end of our sample. Second, there has been concern about the expansion of strategic reserves in non-OECD countries such as China. Non-OECD strategic reserves are not included in our inventory data set. However, most of the new Chinese oil storage facilities had yet to be filled by the end of our sample, so our analysis is not likely to be affected much. Thus there is reason to believe that our inventory data, while less than perfect, are still informative. There are several ways of testing this premise.

First, if there were additional information in the oil futures spread that is not already contained in our inventory proxy, rendering our VAR model informationally misspecified, the oil futures spread should Granger cause the remaining model variables (see Giannone and Reichlin 2006). We formally tested this proposition and were unable to reject the null of no Granger causality at conventional significance levels for maturities of 1, 3, 6, 9, and 12 months, consistent with the view that the inventory data are as accurate as the oil futures spread.[11]

A second test is based on extraneous information about the time periods during which speculation is known to have taken place. For example, Yergin (1992) and other oil market historians have described a speculative frenzy in oil markets in the second half of 1979 with heavy inventory buying. This provides an opportunity to test whether our structural model estimates correctly identify this episode. Similarly, we have a strong presumption that speculation mattered in 1986, 1990 and 2002. In section 4, we will show that our inventory data in all these cases generates results consistent with conventional wisdom.

Finally, Baumeister and Kilian (2012a) show that reduced-form VAR models based on our oil inventory proxy generate more accurate real-time out-of-sample forecasts of the real price

---

[11] Similar results also hold for ex-ante real interest rates.

of oil than other methods even during 2009-2011, further strengthening our claim that the oil inventory proxy is reasonably accurate.

3. Identification

An important question is how to distinguish speculative demand shocks from flow demand and flow supply shocks in practice. Our structural VAR model is set-identified based on a combination of sign restrictions and bounds on the implied price elasticities of oil demand and oil supply.[12] Some of these restrictions are implied by economic theory, while others can be motivated based on extraneous information. We impose four sets of identifying restrictions, each of which is discussed in turn.

3.1. Impact Sign Restrictions

The sign restrictions on the impact responses of oil production, real activity, the real price of oil and crude oil inventories are summarized in Table III.1. These restrictions follow directly from the economic model outlined in section 2. Implicitly, these restrictions also identify the fourth innovation, which can be thought of as a conglomerate of idiosyncratic oil demand shocks. Given the difficulty of interpreting this residual shock economically, we do not report results for this fourth shock, but merely note that it is not an important determinant of the real price of oil.

Sign restrictions alone are typically too weak to be informative about the effects of oil demand and oil supply shocks. As demonstrated in Kilian and Murphy (2012) in the context of a simpler model, it is essential that we impose all credible identifying restrictions for identification for the estimates to be economically meaningful. One such set of restrictions relates to bounds on impact price elasticities of oil demand and oil supply.

3.2. Bound on the Impact Price Elasticity of Oil Supply

The price elasticity of oil supply depends on the slope of the oil supply curve. A vertical short-run oil supply curve would imply a price elasticity of zero, for example. An estimate of the impact price elasticity of oil supply may be constructed from the dynamic simultaneous equation model (1) by evaluating the ratio of the impact responses of oil production and of the real price of oil to an unexpected increase in flow demand or in speculative demand. There is a consensus in the literature that this short-run price elasticity of oil supply is close to zero, if not effectively

---

[12] The use of sign restrictions in oil market VAR models was pioneered by Baumeister and Peersman (2012) and refined by Kilian and Murphy (2012). For a general exposition also see Fry and Pagan (2011) and Inoue and Kilian (2011).

zero.[13] This fact suggests the need for an upper bound on this elasticity in selecting the admissible models that allows for steep, but not quite vertical short-run oil supply curves (see Kilian and Murphy 2012). It is important to stress that this additional identifying restriction does not constrain the levels of the impact responses, but merely imposes a bound on their relative magnitude. In our baseline model, we impose a fairly stringent bound of 0.025 on the impact price elasticity of oil supply. Because any such bound is suggestive only, we also experimented with higher bounds. It can be shown that doubling this bound, while increasing the number of admissible models, has little effect on the shape of the posterior distribution of the impulse responses. Even for a bound of 0.1 the 68% quantiles of the posterior distribution of the impulse responses remain qualitatively similar to the baseline model. Moreover, the estimates of the posterior median price elasticity of oil demand reported in section 5 are remarkably robust to this change.

3.3. Bound on the Impact Price Elasticity of Oil Demand

A preliminary estimate of the impact price elasticity of oil demand may be constructed from the estimated model (1) by evaluating the ratio of the impact responses of oil production and of the real price of oil to an unexpected oil supply disruption. This *oil demand elasticity in production* corresponds to the standard definition of the oil demand elasticity used in the literature. It equates the production of oil with the consumption of oil. In the presence of changes in oil inventories that definition is inappropriate. The relevant quantity measure is instead the sum of the flow of oil production and the depletion of oil inventories triggered by an oil supply shock. To our knowledge, this distinction - with the exception of Considine (1997) - has not been discussed in the literature nor has there been any attempt in the literature to estimate this *oil demand elasticity in use.* The reader is referred to Appendix III.A for a formal discussion of how this elasticity can be derived from the structural VAR model.

A natural additional identifying assumption is that the impact elasticity of oil demand in use, η_t^{Use}, must be weakly negative on average over the sample. In addition to bounding the

---

[13] For example, Hamilton (2009b, p. 25) observes that "in the absence of significant excess production capacity, the short-run price elasticity of oil supply is very low." In practice, it often will take years for significant production increases. Kilian (2009a) makes the case that even in the presence of spare capacity, the response of oil supply within the month to price signals will be negligible because changing oil production is costly. Kellogg (2011) using monthly well-level oil production data from Texas finds essentially no response of oil production to either the spot price or the oil futures price.

demand elasticity in use at zero from above, we also impose a lower bound.[14] It is reasonable to presume that the impact price elasticity of oil demand is lower than the corresponding long-run price elasticity of oil demand (see, e.g., Sweeney 1984). A benchmark for that long-run elasticity is provided by studies of nonparametric gasoline demand functions based on U.S. household survey data such as Hausman and Newey (1995) which have consistently produced long-run price elasticity estimates near -0.8. Their estimate suggests a bound of $-0.8 \leq \eta_t^{Use} \leq 0$.[15]

3.4. Dynamic Sign Restrictions

Our final set of restrictions relates to the dynamic responses to a flow supply shock. We impose the additional restriction that the response of the real price of oil to a negative flow supply shock must be positive for at least twelve months, starting in the impact period. This restriction is necessary to rule out structural models in which unanticipated flow supply disruptions cause a decline in the real price of oil below its starting level. Such a decline would be at odds with conventional views of the effects of unanticipated oil supply disruptions. Because the positive response of the real price of oil tends to be accompanied by a persistently negative response of oil production, once we impose this additional dynamic sign restriction, it furthermore must be the case that global real activity responds negatively to oil supply shocks. This is the only way for the oil market to experience higher prices and lower quantities in practice, because in the data the decline of inventories triggered by an oil supply disruption is much smaller than the shortfall of oil production. This implies a joint set of sign restrictions such that the responses of oil production and global real activity to an unanticipated flow supply disruption are negative for the first twelve months, while the response of the real price of oil is positive.

In contrast, we do not impose any dynamic sign restrictions on the responses to oil demand shocks. In particular, we do not impose any dynamic sign restrictions on the responses of global real activity and oil production to speculative oil demand shocks. Given that this shock is a composite of expectations shocks related to shifts in uncertainty and to the anticipation of rising oil demand and/or falling oil supplies, it is not possible to determine the sign of these responses ex ante beyond the impact period.

---

[14] Note that we do not need to restrict the oil demand elasticity in production. Our impact sign restrictions ensure that this elasticity is weakly negative on impact.

[15] In related work, Yatchew and No (2001) using more detailed Canadian data arrive at a long-run gasoline demand elasticity estimate of -0.9, very close to Hausman and Newey's original estimate.

3.5. Implementation of the Identification Procedure

Given the set of identifying restrictions and consistent estimates of the reduced-form VAR model, the construction of the set of admissible structural models follows the standard approach in the literature on VAR models identified based on sign restrictions (see, e.g., Canova and De Nicolo 2002; Uhlig 2005). Consider the reduced-form VAR model $A(L)y_t = e_t$, where y_t is the N-dimensional vector of variables, $A(L)$ is a finite-order autoregressive lag polynomial, and e_t is the vector of white noise reduced-form innovations with variance-covariance matrix Σ_e. Let ε_t denote the corresponding structural VAR model innovations. The construction of structural impulse response functions requires an estimate of the $N \times N$ matrix $\tilde{B} \equiv B_0^{-1}$ in $e_t = \tilde{B}\varepsilon_t$. Let $\Sigma_e = P\Lambda P'$ and $B = P\Lambda^{0.5}$ such that B satisfies $\Sigma_e = BB'$. Then $\tilde{B} = BD$ also satisfies $\tilde{B}\tilde{B}' = \Sigma_e$ for any orthogonal $N \times N$ matrix D. One can examine a wide range of possibilities for $\tilde{B}$ by repeatedly drawing at random from the set $\mathbf{D}$ of orthogonal matrices D. Following Rubio-Ramirez, Waggoner and Zha (2010) we construct the set $\tilde{\mathbf{B}}$ of admissible models by drawing from the set $\mathbf{D}$ and discarding candidate solutions for $\tilde{B}$ that do not satisfy a set of a priori restrictions on the implied impulse response functions. In practice, this procedure may be implemented conditional on the conventional maximum likelihood/least squares estimator of $A(L)$ and Σ_e in the reduced-form VAR model. This allows the resulting impulse response estimates to be given a frequentist interpretation. To summarize, the procedure consists of the following steps:

1) Draw an $N \times N$ matrix K of $NID(0,1)$ random variables. Derive the QR decomposition of K such that $K = Q \cdot R$ and $QQ' = I_N$.
2) Let $D = Q'$. Compute impulse responses using the orthogonalization $\tilde{B} = BD$. If all implied impulse response functions satisfy the identifying restrictions, retain D. Otherwise discard D.
3) Repeat the first two steps a large number of times, recording each D that satisfies the restrictions and record the corresponding impulse response functions.

The resulting set $\tilde{\mathbf{B}}$ corresponds to the set of all admissible structural VAR models.

The estimation uncertainty underlying these structural impulse response estimates may be assessed by frequentist or Bayesian methods. We adopt the latter approach and follow the standard approach in the literature of specifying a diffuse Gaussian-inverse Wishart prior distribution for the reduced-form parameters and a Haar distribution for the rotation matrix (see, e.g., Inoue and Kilian 2011). The posterior distribution of the structural impulse responses is obtained by applying our identification procedure to each draw of $A(L)$ and Σ_e from their posterior distribution.

4. Estimation Results

The identifying restrictions described in section 3 do not yield point-identified structural impulse responses, but a range of models consistent with the identifying assumptions. For expository purposes, in the analysis below we focus on one model among the admissible structural models obtained conditional on the least squares estimate of the reduced-form. The results shown are for the model that yields an impact price elasticity of oil demand in use closest to the posterior median of this elasticity among the candidate models that satisfy all identifying restrictions. We also conducted the same analysis with every other admissible structural model and verified that our main results are robust to the choice of admissible model. The only difference is that some admissible models assign even more explanatory power to flow demand shocks than the benchmark model at the expense of speculative demand shocks.

4.1. Responses to Oil Supply and Oil Demand Shocks

Figure III.1 plots the responses of each variable in this benchmark model to the three oil supply and oil demand shocks along with the corresponding pointwise 68% posterior error bands. All shocks have been normalized such that they imply an increase in the real price of oil. In particular, the flow supply shock refers to an unanticipated oil supply disruption. Figure III.1 illustrates that the role of storage differs depending on the nature of the shock. A flow supply disruption causes inventories to be drawn down in an effort to smooth production of refined products. A positive flow demand shock causes almost no change in oil inventories on impact, followed by a temporary drawdown of oil inventories. After one year, oil inventories reach a level in excess of their starting level. A positive speculative demand shock causes a persistent increase in oil inventories.

A negative flow supply shock is also associated with a reduction in global real activity and a persistent drop in oil production, but much of the initial drop is reversed within the first half year. The real price of oil rises only temporarily. It peaks after three months. After one year, the real price of oil falls below its starting value, as global real activity drops further. A positive shock to the flow demand for crude oil, in contrast, is associated with a persistent increase in global real activity. It causes a persistent hump-shaped increase in the real price of oil with a peak after one year. Oil production also rises somewhat, but only temporarily. Finally, a positive speculative demand shock is associated with an immediate jump in the real price of oil. The real price response overshoots, before declining gradually. The effects on global real activity and global oil production are largely negative, but small. These estimates imply a larger role for flow supply shocks than the structural VAR model in Kilian (2009a,b), for example, illustrating the importance of explicitly modeling speculative demand shocks and oil inventories.

4.2. What Drives Fluctuations in Oil Inventories and in the Real Price of Oil?

It can be shown that in the short run, 29% of the variation in crude oil inventories is driven by speculative demand shocks, followed by oil supply shocks with 26%. Flow demand shocks have a negligible impact with 2%. At long horizons, in contrast, the explanatory power of speculative demand shocks declines to 27% and that of flow supply shocks to 24%, while the explanatory power of flow demand shocks increases to 15%. This evidence suggests that, on average, fluctuations in oil inventories mainly reflect speculative trading as well as production smoothing by refiners in response to oil supply shocks. This contrasts with a much larger role of flow demand shocks in explaining the variability of the real price of oil. For example, in the long run, 87% of the variation in the real price of oil can be attributed to flow demand shocks, compared with 9% due to speculative demand shocks and 3% due to flow supply shocks.

Impulse responses and forecast error variance decompositions are useful in studying average behavior. To understand the historical evolution of the real price of oil, especially following major exogenous events in oil markets, it is more useful to compute the cumulative effect of each shock on the real price of oil and on the change in oil inventories. Figure III.2 allows us to assess the quantitative importance of speculative demand shocks as opposed to other demand and supply shocks at each point in time since the late 1970s.[16]

4.3. Did Speculators Cause the Oil Price Shock of 2003-2008?

A common view in the literature is that speculators caused part or all of the run-up in the real price of oil between 2003 and mid-2008. Especially the sharp increase in the real price of oil in

---

[16] We do not include the contribution of the residual demand shock because that shock makes no large systematic contribution to the evolution of the real price of oil.

2007/08 has been attributed to speculation. The standard interpretation is (a) that there was an influx of financial investors into the oil futures market, (b) that this influx drove up oil futures prices, and (c) that the increase in oil futures prices was viewed by spot market participants as a signal of an expected increase in the price of oil, shifting inventory demand and hence causing the real spot price to increase.

This explanation implies that speculative demand shocks in the structural model should explain the bulk of the surge in the real price of oil after 2003. Figure 2 shows that there is no evidence to support this view. There has been no systematic upward movement in the real price of oil after 2003 associated with speculative demand shocks. This result has far-reaching implications. First, in the policy debate, it is common to distinguish between normal speculation in oil markets that reflects expected fundamentals and purely financial speculation that is viewed as excessive. While there is no operational definition of "excessive speculation", as discussed in Fattouh et al. (2012), the evidence in Figure III.2 suggests that the distinction between normal speculation and speculation that is excessive is moot, for if there is no speculation at all, there cannot be excessive speculation under any definition.

Second, this result tells us that an exogenous influx of financial speculators cannot have driven up the oil futures price, because – under the standard assumption of arbitrage between the futures and spot markets for oil maintained by proponents of the financial speculation hypothesis – the absence of speculation in the spot market also rules out an exogenous influx of speculators in oil futures markets.[17]

A competing view of speculation is that OPEC in anticipation of even higher oil prices held back its production after 2001, using oil below ground effectively as inventories (see, e.g., Hamilton 2009a, p. 239). One way of testing this hypothesis is through the lens of our structural model. In our model, OPEC holding back oil production in anticipation of rising oil prices would be observationally equivalent to a negative flow supply shock. Figure III.2 provides no indication that negative flow supply shocks were an important determinant of the real price of oil between 2003 and mid-2008. What evidence there is of a small supply-side driven increase in the

---

[17] Our analysis of the spot market would remain valid if this arbitrage were impeded or broke down completely, but the oil futures price would become disconnected from the spot price. In the limiting case of no arbitrage we would be unable to infer from our model whether there is speculation in oil futures markets, although we could still infer whether there is speculation taking place in the spot market. Clearly, a situation in which arbitrage breaks down is not consistent with the scenario envisioned by researchers who attribute rising spot prices to speculation by financial investors, because in that case speculation-driven increases in the oil futures price could not possibly be transmitted to the spot price of oil.

real price of oil is dwarfed by the price increases associated with flow demand shocks. Hence, we can reject the speculative supply shock hypothesis.

An alternative explanation of the evolution of the real price of oil is the peak oil hypothesis, which predicted that around 2006 world oil production should have peaked. If so, one would have expected to see a sequence of negative flow supply shocks drive up the real price of oil after 2005, but we already showed that flow supply shocks have very limited explanatory power. The peak oil hypothesis could also affect the real price of oil if traders rightly or wrongly believed in this hypothesis and stocked up on oil in anticipation of a shortage of oil. That explanation would be observationally equivalent to the financial speculation hypothesis, however, which we already rejected on the basis of the results in Figure III.2. Hence, peak oil may be safely ruled out as an explanation of the surge in the real price of oil after 2003, along with financial speculation and speculation by oil producers.

Instead our model provides a different explanation. It supports the substantive conclusion in Kilian (2009a,b) that the surge in the real price of oil between 2003 and mid-2008 was mainly caused by shifts in the flow demand for crude oil associated with the global business cycle. It is important to stress that this result does not arise by construction. Indeed, the identification of our model is quite different from that in Kilian (2009a), and, as we will show below, the two models may produce substantively different empirical results. Our model shows that the run-up in the real price of oil occurred because of the cumulative effects of many positive flow demand shocks over the course of several years. It may seem unlikely ex ante that a model would generate many more positive than negative demand shocks between 2003 and mid-2008, but Kilian and Hicks (2012) show that this feature is consistent with the errors in professional real GDP forecasts during this period. Even professional forecasters persistently underestimated global growth during 2003-08, especially growth in emerging Asia, lending credence to our model results.

This situation only changed with the financial crisis of late 2008. The V-shaped dip in the real price of oil in 2008/09 coincided with a similar dip in the global real activity measure and is largely driven by flow demand shocks. A similar, if much less pronounced, dip had followed the Asian crisis in 1997. Whereas the recovery from the all-time low in the real price of oil in 1999-2000 resulted from a combination of coordinated OPEC oil supply cuts, a gradual increase in flow demand (often associated with the U.S. productivity boom) and increased speculative demand perhaps in anticipation of increased future real activity and/or further oil supply reductions, the resurgence of the real price of oil starting in early 2009 reflected primarily increased flow demand (see Figure III.2).

We conclude that economic fundamentals on the demand side of the oil market are capable of explaining the evolution of the real price of oil during the last decade. No non-standard explanations are required. This finding is important because it implies that further regulation of oil markets would have done nothing to stem the increase in the real price of oil. Indeed it shows that there is no basis for the premise that such regulation is required to lower the real price of oil. Our structural model also implies that even dramatic increases in U.S. oil production would not lower the real price of oil substantially at the global level, while a full recovery of the global economy would raise the real price of oil by as much as 50 dollars in real terms (see Baumeister and Kilian 2012b).

4.4. The Inventory Puzzle of 1990

Although speculative motives played no important role after 2003, there are other oil price shock episodes when they did, suggesting that our model has the ability to detect speculative demand shocks when they exist. One particularly interesting example is the oil price spike associated with the Persian Gulf War of 1990/91. In related work, Kilian (2009a) presented evidence based on a model without oil inventories that the 1990 oil price increase was driven mainly by a shift in speculative demand (reflecting concerns about future oil supplies from neighboring Saudi Arabia) rather than the physical reduction in oil supplies associated with the war. As noted by Hamilton (2009a), this result is puzzling upon reflection because oil inventories moved little and, if anything, slightly declined following the invasion of Kuwait. This observation prompted Hamilton to reject the hypothesis that shifts in speculative demand were behind the sharp increase in the real price of oil in mid-1990 and its fall after late 1990. Given the consensus that flow demand did not move sharply in mid-1990, Hamilton suggested that perhaps this price increase must be attributed to flow supply shocks after all. The inventory data, however, seem just as inconsistent with this alternative hypothesis. Inventories declined in August of 1990, but only by one third of a standard deviation of the change in inventories. Given one of the largest unexpected oil supply disruptions in history, one would have expected a much larger decline in oil inventories given the impulse response estimates in Figure III.1. Moreover, there is general agreement that flow supply shocks cannot explain the collapse of the real price of oil in late

1990. In light of this evidence, neither the supply shock explanation nor the speculative demand shock explanation by itself seems compelling.

Our econometric model resolves this inventory puzzle. The explanation is that the invasion of Kuwait in August of 1990 represented two shocks that occurred simultaneously. On the one hand it involved an unexpected flow supply disruption and on the other an unexpected increase in speculative demand. Whereas the flow supply shock caused a decline in oil inventories, increased speculative demand in August caused an increase in oil inventories, with the net effect being a modest decline in oil inventories. At the same time, the observed large increase in the real price of oil was caused by both shocks working in the same direction.

The historical decomposition in Figure III.3 contrasts the price and inventory movements caused by each shock during 1990/91. It shows that about one third of the price increase from July to August of 1990 was caused by speculative demand shocks and two thirds by flow supply shocks. This result is in sharp contrast to the estimates in Kilian (2009a) who found no evidence of oil supply shocks contributing to this increase, illustrating again that the inclusion of inventories in the structural model matters.

Figure III.3 also highlights that the decline in the real price of oil starting in November of 1990, when the threat of Saudi oil fields being captured by Iraq had been removed by the presence of U.S. troops, was almost entirely caused by a reduction in speculative demand rather than increased oil supplies. The latter observation is consistent with evidence in Kilian (2008a) that it is difficult to reconcile the sharp decline in the real price of oil starting in late 1990 with data on oil production. The evidence of a sharp decline in speculative demand in late 1990 in turn raises the obvious question of when and why speculative demand had surged in the first place. The bottom panel of Figure III.2 reveals there was in fact a substantial increase in speculative demand already in the months leading up to the invasion. This result is consistent with a sharp increase in oil inventories in the months leading up to the invasion. One interpretation is that the invasion was anticipated by informed oil traders or, more likely, that traders responded to evidence of increased political tension in the Middle East.[18] The reason that this increase in the speculative component of the real price of oil went unnoticed by the general public was a simultaneous substantial unexpected increase in oil production with offsetting effects on the real price of oil in early 1990, as shown in the top panel of Figure III.2. In fact, that

---

[18] In this regard, Gause (2002) notes a shift in Iraqi foreign policy toward a more aggressive stance in early 1990.

expansion of oil production served as the motivation for Iraq's increasing hostility to neighboring countries such as Kuwait which it accused of undermining the price of oil, making it more difficult for Iraq to service its foreign debt. Taken in conjunction our evidence implies a much larger role for speculative demand in 1990/91 than the data for the month of August alone would suggest.

4.5. What Caused the 1979 and 1980 Oil Price Shocks?

Speculative demand also played an important role in 1979. The traditional interpretation of this episode is that this oil price increase was caused by flow supply disruptions associated with the Iranian Revolution of late 1978 and early 1979. Much of the observed increase in the real price of oil, however, only occurred later in 1979 after Iranian oil production had resumed. Barsky and Kilian (2002) therefore attribute the price increase starting in May of 1979 and extending into 1980 in part to increased flow demand for oil and in part to a substantial increase in speculative demand for oil, consistent with anecdotal evidence from oil industry sources and with the perception of a noticeable increase in the risk of an oil supply disruption in the Persian Gulf in 1979, coupled with expectations of strong flow demand.[19]

This hypothesis is testable in our model. Figure III.2 shows that not only was there a dramatic and persistent increase in the real price of oil driven by positive flow demand shocks in 1979 and 1980 (not unlike the persistent price increase after 2003), but that increase was reinforced after May of 1979 by a sharp increase in speculative demand, exactly as described by Yergin (1992). Whereas flow demand pressures on the real price of oil gradually receded starting in 1981, speculative demand pressures on average remained relatively high until the collapse of OPEC in late 1985. In contrast, there is little evidence of flow supply shocks being responsible for the oil price surge of 1979, consistent with the fact that overall global oil production increased in 1979, reflecting additional oil production outside of Iran. Only in late 1980 and early 1981 is there a moderate spike in the real price of oil driven by flow supply shocks, in part associated with the outbreak of the Iran-Iraq War (see Figure III.2).

---

[19] For example, Terzian (1985, p. 260) writes that in 1979 "spot deals became more and more infrequent. The independent refineries, with no access to direct supply from producers, began to look desperately for oil on the so-called 'free market'. But from the beginning of November, most of the big oil companies invoked *force majeure* and reduced their oil deliveries to third parties by 10% to 30%, when they did not cut them off altogether. Everybody was anxious to hang on to as much of their own oil as possible, until the situation had become clearer. The shortage was purely psychological, or 'precautionary' as one dealer put it." Also see Yergin (1992, p. 687).

It is useful to explore the price and inventory dynamics in 1979 in more detail. The historical decomposition in Figure III.4 confirms that negative flow supply shocks caused a temporary drop in oil inventories in December of 1978 and January of 1979, but for the next half year positive flow supply shocks increased oil inventories. This result is also consistent with the fact that global oil production starting in April of 1979 exceeded its level prior to the Iranian Revolution. At the same time, after March of 1979, repeated speculative demand shocks caused a persistent accumulation of inventories, while driving up the real price of oil. The inventory accumulation continued into 1980.Thus, there is no indication that flow supply shocks played an important role in the oil price surge of 1979.

It was not until September of 1980 when the Iran-Iraq War broke out that the oil market experienced another major disruption of flow supply. This event is once again associated with declining oil inventories initially and subsequently rising inventories driven by unexpected flow supply increases, reflecting in part the growing importance of new non-OPEC oil producers. As Figure III.5 shows, the increase in the real price of oil in response to this flow supply disruption was slightly larger than the price response to the 1979 supply disruption. There is also evidence of a small resurgence of speculative demand following the outbreak of the war, reflected in rising inventories and a higher real price of oil.

4.6. The Collapse of OPEC in 1986

In late 1985, Saudi Arabia decided that it would no longer attempt to prop up the price of oil by reducing its oil production, creating a major positive shock to the flow supply of oil. The same event also markedly changed market perceptions about OPEC's market power. Figure III.6 shows that, as expected, the positive flow supply shock in early 1986 drove down the real price of oil, while oil inventories rose in response. Simultaneously, a drop in speculative demand reinforced the decline in the real price of oil, while lowering inventory holdings. This pattern is similar to the pattern in Figure III.3, except in reverse. Although OPEC attempted to reunite and control production in 1987, amidst increased speculation, these attempts proved unsuccessful in the long run.

4.7. The Venezuelan Crisis and Iraq War of 2002/03

Figure III.7 focuses on the flow supply shock of 2002/2003 when within months first Venezuelan oil production slowed considerably at the end of 2002 and then Iraqi oil production ceased altogether in early 2003. The combined cutback in oil production was of a magnitude similar to the oil supply disruptions of the 1970s (see Kilian 2008a). Figure III.7 shows that this event reflected a combination of negative flow supply shocks and positive speculative demand shocks.

The Venezuelan oil supply crisis of late 2002 was associated with declining oil inventories, consistent with an unexpected oil supply disruption, but this period was also coincided with an increase in speculative demand in anticipation of the 2003 Iraq War that dampened the decline in inventories, while reinforcing the increase in the real price of oil. The military conflict in Iraq lasted from late March 2003 until the end of April 2003. Despite the additional loss of Iraqi output in early 2003, global oil production unexpectedly increased. The production shortfalls in Iraq and Venezuela were more than offset at the global level by increased oil production elsewhere. These positive flow supply shocks lowered the real price of oil starting in early 2003 and resulted in positive inventory accumulation. At the same time, as early as March 2003, lower speculative demand caused the real price of oil to drop and oil inventories to fall. Again the effect of the two shocks on inventories was offsetting, whereas the effect on the price worked in the same direction. This last example again underscores that geopolitical events in the Middle East matter not merely because of the disruptions of the flow supply of oil they may create, but also because of their effect on speculative demand.

5. Implications of the Model for the Short-Run Price Elasticity of Oil Demand

The short-run price elasticity of oil demand has important implications for theoretical models of speculative demand. For example, it is a key parameter in the theoretical models of speculation discussed in Caballero, Farhi and Gourinchas (2008) and Hamilton (2009a). All else equal, standard models of speculation imply that oil inventories must increase to enable the real price of oil to increase. Recently, Hamilton (2009a) observed that speculation in oil futures markets may drive up the real price of oil without any increase in oil inventories if refiners are able to pass on fully to gasoline consumers an exogenous increase in the real price of oil driven by speculation. This result requires the demand for gasoline to be completely price-inelastic. As shown in Appendix III.B, the short-run price elasticity of gasoline demand is approximately of the same magnitude as the short-run price elasticity of oil demand in use. Whether the limiting case discussed in Hamilton (2009a) is empirically relevant thus depends on the magnitude of the short-run price elasticity of oil demand.

While there is little doubt that the price elasticity of oil supply is near zero in the short run, the literature does not offer much direct evidence on the magnitude of the short-run price elasticity of oil demand. It is widely believed that this elasticity is close to zero, making it difficult to rule out the limiting case described by Hamilton. Although there is no shortage of elasticity estimates in the literature that seem to confirm this impression, these estimates suffer from two limitations.

First, much of the existing literature has attempted to estimate the oil demand elasticity from dynamic reduced-form models that do not distinguish between oil demand and oil supply shocks (see, e.g., Dahl 1993; Cooper 2003). This is not possible because the identification of the demand elasticity requires an exogenous shift of the contemporaneous oil supply curve along the contemporaneous oil demand curve within the context of a structural model. Reduced-form estimates of the oil demand elasticity fail to account for the endogeneity of the price of crude oil and hence are biased toward zero. This fact helps explain the low elasticity estimates typically reported in the literature.[20]

Second, typical estimates of the oil demand elasticity in the literature have been based on models that equate the percent change in quantity with the percent change in the production of crude oil, ignoring the existence of oil inventories. In this paper, we refer to this conventional elasticity measure as the oil demand elasticity in production, denoted by $\eta^{O,\Pr oduction}$. A more appropriate definition of the price elasticity of oil demand for policy questions is the elasticity in use. The latter demand elasticity is based on the change in the use of oil, defined as the sum of the change in oil production and of the depletion of oil inventories, which more accurately captures the response of oil consumers.

5.1. The Short-Run Price Elasticity of Oil Demand in Production

Our structural model of the oil market may be used to obtain direct estimates of the short-run price elasticity of oil demand in production and in use, allowing us to the empirical relevance of models relying on a zero price elasticity of oil demand. The elasticity in production can be estimated from model (1) as the ratio of the impact response of oil production to a flow supply shock relative to the impact response of the real price of oil. Our posterior median estimate of this elasticity, as shown in the first column of the upper panel of Table III.2, is -0.44. This

---

[20] Producers of these estimates sometimes acknowledge the need for instrumental variable estimation methods, but, having acknowledged this point, tend to revert to using OLS, given the absence of suitable instruments.

estimate is seven times higher than typical conjectures in the recent literature. It is also much higher in magnitude than conventional reduced-form regression estimates of this elasticity. For example, surveys by Dahl (1993) and Cooper (2003) report estimates between -0.05 and -0.07. The difference in results can be attributed to the difference between estimating a structural and a reduced-form model.[21] The first column in the upper panel of Table III.2 also shows the 68% posterior credible set for this elasticity. The model assigns substantial probability mass to values between -0.80 and -0.23 and very little probability mass to values close to zero.

5.2. The Short-Run Price Elasticity of Oil Demand in Use

The posterior median estimate of $\eta^{O,\Pr oduction}$ in the first column of Table III.2, while instructive when compared to conventional estimates, is misleading in that it ignores the role of inventories. Our model also permits the estimation of the price elasticity of oil demand in use, allowing us to assess the role of changes in inventories (see Appendix III.A). By construction, allowing for inventory responses will tend to lower the magnitude of the price elasticity of oil demand. The second column in the upper panel of Table III.2 shows that posterior median estimate of $\eta^{O,Use}$ is only -0.26 compared with the estimate of -0.44 for the demand elasticity in oil production. While this large reduction in the magnitude of the elasticity highlights the importance of accounting for changes in oil inventories, the revised median estimate is still four times larger than conventional elasticity estimates. Indeed, this result illustrates that relatively high short-run price elasticities of oil demand are fully compatible with the view that economic fundamentals are responsible for the surge in the real price of oil after 2003.[22] Moreover, the second column of Table III.2 again shows that there is little probability mass on elasticity values close to zero, casting doubt on models of speculation that do not involve a change in oil inventories.[23]

---

[21] One way of demonstrating this point is to note that fitting the conventional reduced-form log-level specification used in some of the earlier literature to our data would yield an elasticity estimate of only -0.02 in line with the existing consensus. Another way of putting these results into perspective is to observe that other recent studies relying on alternative structural models have obtained similarly large oil demand elasticity estimates ranging from -0.35 to -0.41 (see, e.g., Serletis et al. 2010; Baumeister and Peersman 2012; Bodenstein and Guerrieri 2011).

[22] The lower panels of Table 2 show that these estimates are quite robust to relaxing the upper bound on the impact price elasticity of oil supply. Relaxing this bound to 0.050 or even to 0.100 raises the median oil demand elasticity in use slightly without affecting the substance of the conclusions.

[23] It is worth noting that our estimate of the price elasticity of gasoline demand is larger than some estimates in the literature. For example, Hughes, Knittel and Sperling (2008) report an average elasticity estimate of -0.18 based on U.S. data for 1975-80 and for 2001-06, similar to estimates in Dahl and Sterner (1991). On the other hand, Burger and Kaffine (2009) report estimates as high as -0.29. Our estimate also is smaller than the instrumental variable regression estimate of the gasoline tax elasticity of gasoline demand of -0.47 reported in Davis and Kilian (2011) with a standard error of 0.23. An unresolved question is to what extent the price elasticity of gasoline demand may

6. Conclusion

Standard structural VAR models of the market for crude oil implicitly equate oil production with oil consumption and ignore the role of oil inventories. Traditionally these models have focused on shocks to the flow supply of oil and the flow demand for oil. In this paper we augmented the structural model to include shocks to inventory demand reflecting shifts in expectations about future oil supply and future oil demand that cannot be captured by flow demand or flow supply shocks. Such speculative demand shocks must be represented as shifts of the contemporaneous oil demand curve rather than the contemporaneous oil supply curve, even if the shift in expectations is about a cut in future oil supplies rather than an increase in future oil demand. The reason is that traders in anticipation of the expected oil shortage will buy and store crude oil now with the expectation of selling later at a profit. We proposed a dynamic simultaneous equation model including crude oil inventories that allows the simultaneous identification of all three types of shocks conditional on past data.

The inclusion of oil inventories matters. For example, our structural model implies a larger role for flow supply shocks in explaining fluctuations in the real price of oil than previous estimates. The added explanatory power of flow supply shocks for the real price of oil, especially in 1990, comes at the expense of the explanatory power of speculative demand shocks. We showed that, nevertheless, the largest and most persistent fluctuations in the real price of oil since the 1970s have been associated primarily with business cycle fluctuations affecting the demand for crude oil. Of particular interest for recent policy debates is the sustained increase in the real price of oil from 2003 until mid-2008. In the context of a model that nests all leading explanations of how this oil price surge came about, we were able to provide direct evidence against the popular view that this increase was driven by speculators. This conclusion holds even for the 2007/08 period. Shifts in speculative demand played a more important role during several earlier oil price shock episodes, however, notably in 1979, 1986, 1990, and 2002. We showed that, without accounting for shifts in the speculative demand for oil, it is not possible to understand the evolution of the real price of oil during these episodes.

Our analysis also suggests that there is no evidence that peak oil or that deliberate production cutbacks by oil producers had much bearing on the recent oil price fluctuations.

---

have declined in magnitude in very recent years and at what time (see, e.g., Hughes, Knittel, and Sperling 2008). It would take a substantial decline, however, to make the limiting case of a zero elasticity discussed in Hamilton (2009a) empirically relevant.

Rather our results support recent findings in the literature that the sustained run-up in the real price of oil between 2003 and mid-2008 was caused primarily by shifts in the global flow demand for oil. Likewise, the collapse of the real price of oil in late 2008 and its partial recovery since early 2009 have been primarily driven by increased flow demand. The model implies that the real price of oil is expected to rise further as the global economy recovers from the financial crisis, creating a policy dilemma, unless energy consumption can be reduced or new energy sources can be found. By contrast, additional regulation of oil derivatives markets is not likely to lower the real price of oil nor can increased domestic oil production in the U.S. be expected to have much of an effect on the real price of oil (also see Baumeister and Kilian 2012b).

Our model also reconciles some seemingly puzzling observations related to earlier oil price surges. For example, it has been noted that, following the outbreak of the Persian Gulf War in August 1990, oil inventories did not increase as one would have expected in response to a positive speculative demand shock. At the same time, the absence of a sharp decline in oil inventories in August of 1990 is inconsistent with the view that the price increase reflected a negative oil supply shock. We demonstrated that this inventory puzzle can be resolved with the help of a structural oil market model. Our analysis showed that the price and inventory data can be explained only based on a combination of these two shocks. Because the implied inventory responses are of opposite sign, the net effect in inventories is close to zero, where the sharp price increase reflects the fact that the implied price responses are of the same sign. Similar relationships were shown to hold during other key historical episodes. These examples illustrate that it is essential to rely on structural models rather than reduced-form evidence in interpreting the price and quantity data.

The use of a structural regression model also is important for the construction of the short-run price elasticity of oil demand. For example, Hamilton (2009a,b) suggested that 1978-81 is one episode where one might clearly and without a regression model attribute cumulative changes in the price of oil to exogenous oil supply shifts only, allowing one to construct a demand elasticity estimate from the ratio of cumulative changes in quantities and prices for that period. The structural model we have analyzed suggests otherwise. We showed that oil demand shocks were the main cause of the observed oil price increase in 1978-81. Oil supply shocks played a small role only, violating the premise of Hamilton’s calculations.

We also observed that traditional estimates of the short-run price elasticity of oil demand are not credible. One problem is that conventional estimates of this elasticity from dynamic reduced-form regressions, as in Dahl (1993) and Cooper (2003), have ignored the endogeneity of the real price of oil, causing the elasticity estimate to be downward biased. Moreover, existing estimates, including the structural estimates recently provided by Baumeister and Peersman (2012), have ignored the role of inventories in smoothing oil consumption in response to oil supply shocks. We provided a model that allows the estimation of both the traditional oil demand elasticity in production and of the more relevant oil demand elasticity in use which incorporates changes in oil inventories. Our short-run elasticity estimates are substantially higher than standard estimates cited in the literature, casting doubt on models of speculative trading based on perfectly price-inelastic short-run demand for oil and for gasoline.

Table III.1: Sign Restrictions on Impact Responses in VAR Model

	Flow supply shock	Flow demand shock	Speculative demand shock
Oil production	-	+	+
Real activity	-	+	-
Real price of oil	+	+	+
Inventories			+

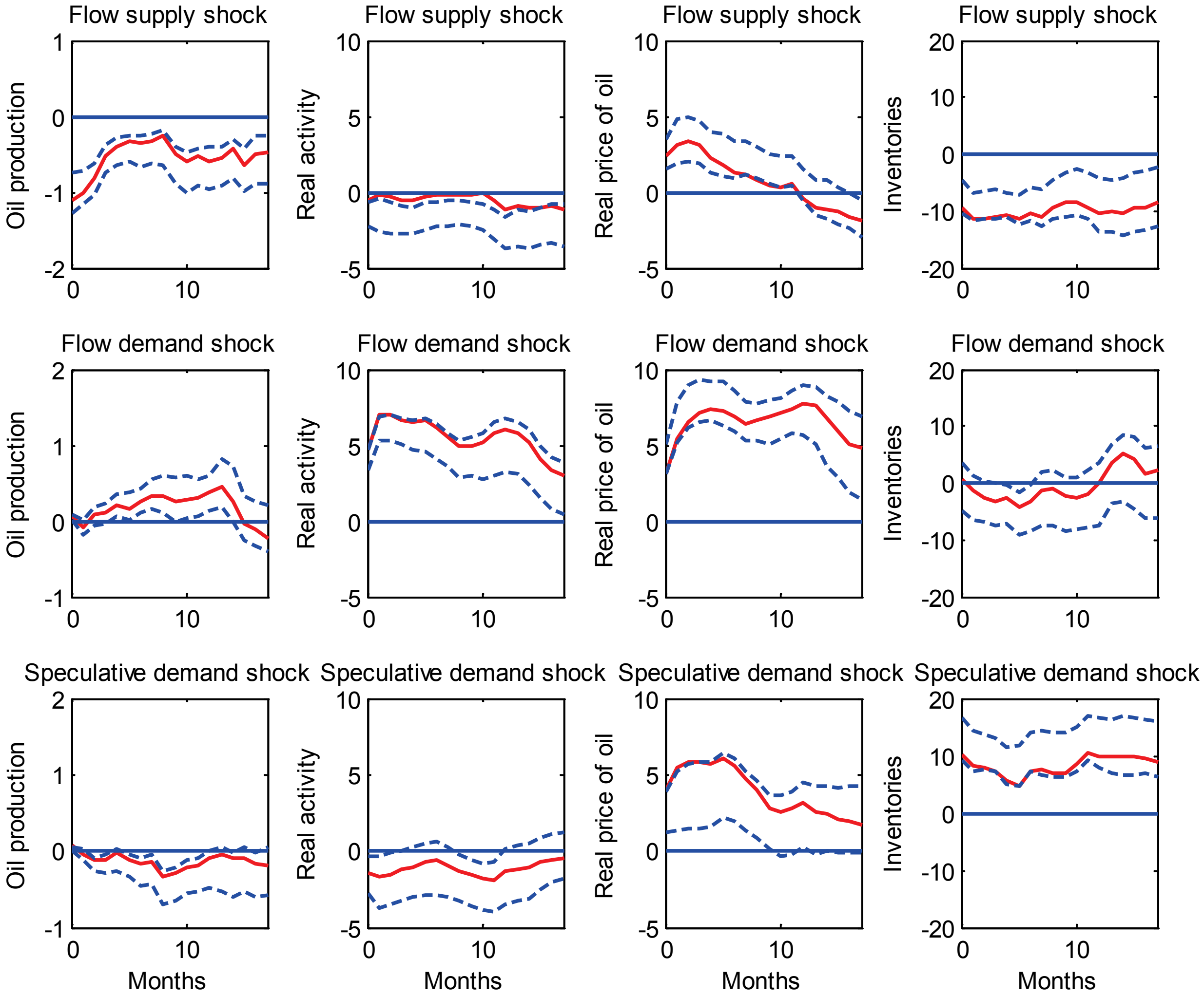
Note: All structural shocks have been normalized to imply an increase in the real price of oil. Missing entries mean that no sign restriction is imposed.

Table III.2: Posterior Distribution of the Short-Run Price Elasticity of Demand for Crude Oil

		$\eta^{O,\mathrm{Pr}oduction}$	$\eta^{O,Use}$
$\eta_t^{Supply} \leq 0.025$	16th Percentile	-0.80	-0.54
	50th Percentile	-0.44	-0.26
	84th Percentile	-0.23	-0.09
$\eta_t^{Supply} \leq 0.050$	16th Percentile	-0.80	-0.57
	50th Percentile	-0.45	-0.27
	84th Percentile	-0.29	-0.09
$\eta_t^{Supply} \leq 0.100$	16th Percentile	-0.76	-0.61
	50th Percentile	-0.47	-0.30
	84th Percentile	-0.24	-0.10

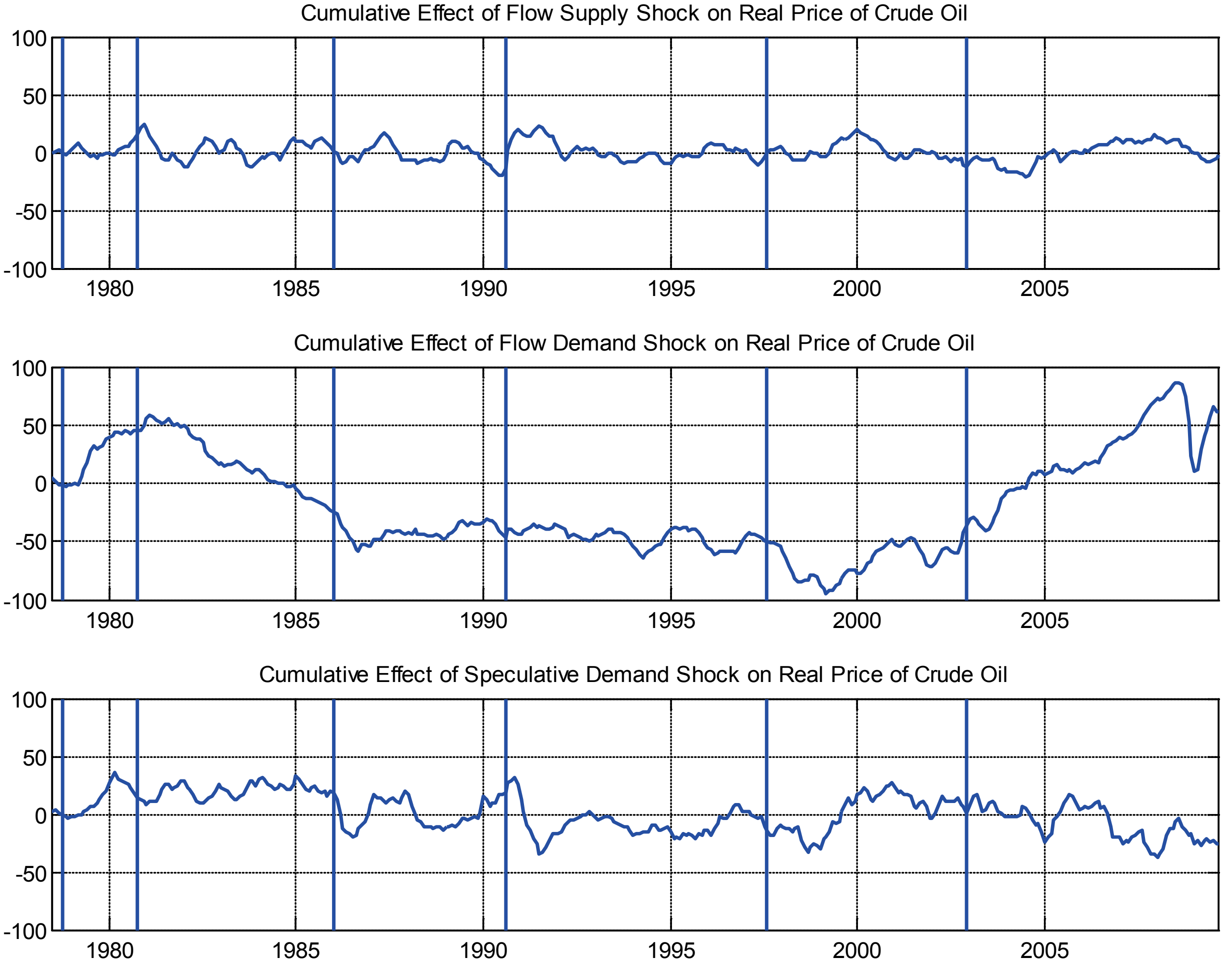
Note: $\eta^{O,\mathrm{Pr}oduction}$ refers to the impact price elasticity of oil demand in production and $\eta^{O,Use}$ to the average impact price elasticity of oil demand in use. The latter definition accounts for the role of inventories in smoothing oil consumption. η^{Supply} refers to the impact price elasticity of oil supply.

Figure III.1: Structural Impulse Responses: 1973.2-2009.8



Note: Solid lines indicate the impulse response estimates for the model with an impact price elasticity of oil demand in use closest to the posterior median of that elasticity among the admissible structural models obtained conditional on the least-squares estimate of the reduced-from VAR model. Dashed lines indicate the corresponding pointwise 68% posterior error bands. Oil production refers to the cumulative percent change in oil production and inventories to cumulative changes in inventories.

Figure III.2: Historical Decompositions for 1978.6-2009.8



Note: Based on benchmark estimate as in Figure III.1. The vertical bars indicate major exogenous events in oil markets, notably the outbreak of the Iranian Revolution in 1978.9 and of the Iran-Iraq War in 1980.9, the collapse of OPEC in 1985.12, the outbreak of the Persian Gulf War in 1990.8, the Asian Financial Crisis of 1997.7, and the Venezuelan crisis in 2002.11, which was followed by the Iraq War in early 2003. In constructing the historical decomposition we discard the first five years of data in an effort to remove the transition dynamics.

Figure III.3: Historical Decompositions for the Persian Gulf War Episode of 1990/91

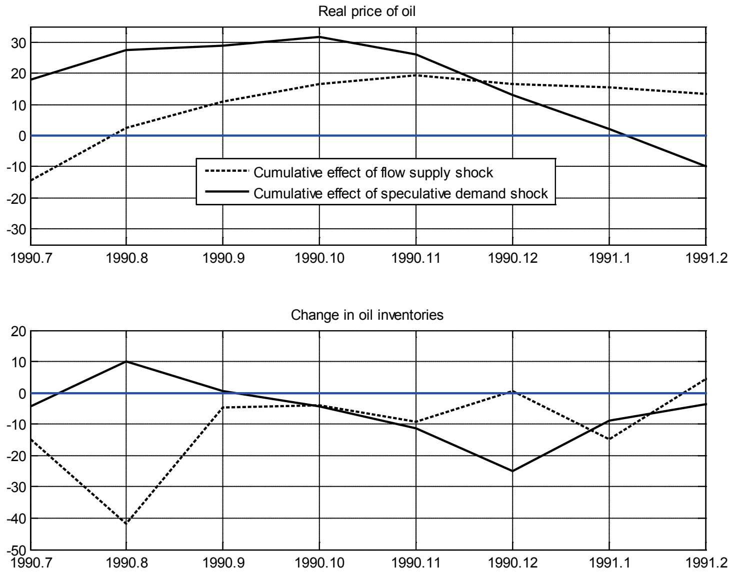


Figure III.4: Historical Decompositions for the Iranian Revolution of 1978/79

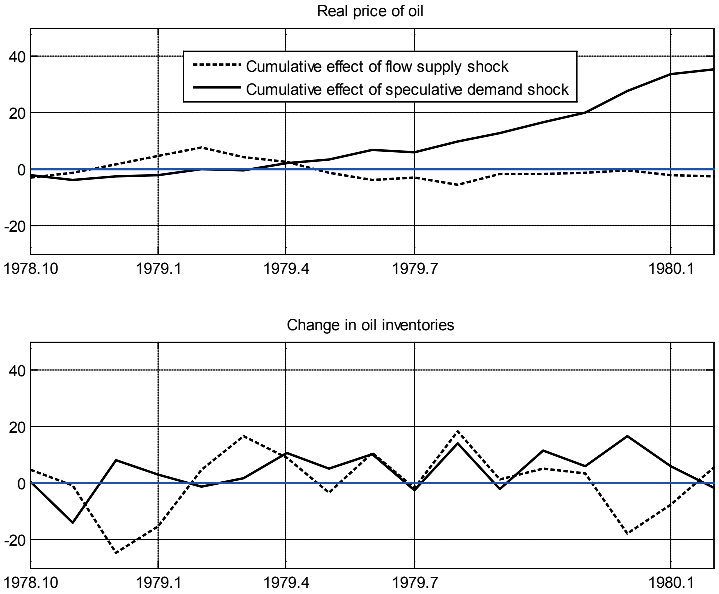


Figure III.5: Historical Decompositions for the Outbreak of the Iran-Iraq War in 1980

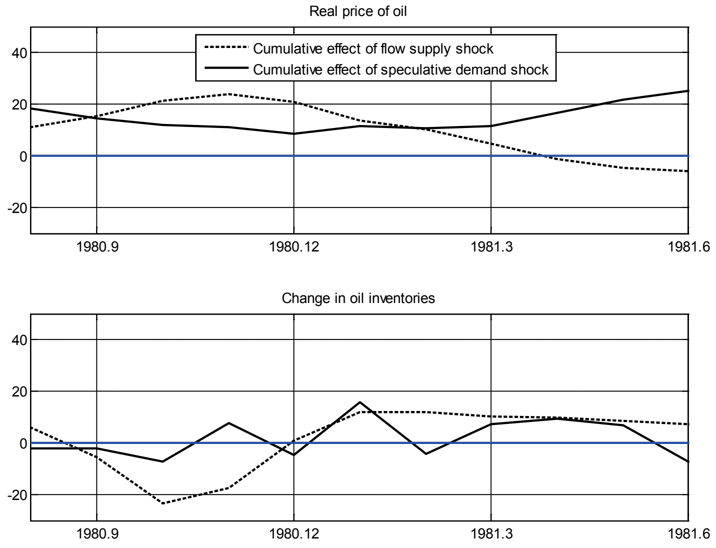


Figure III.6: Historical Decompositions for the Collapse of OPEC in 1986

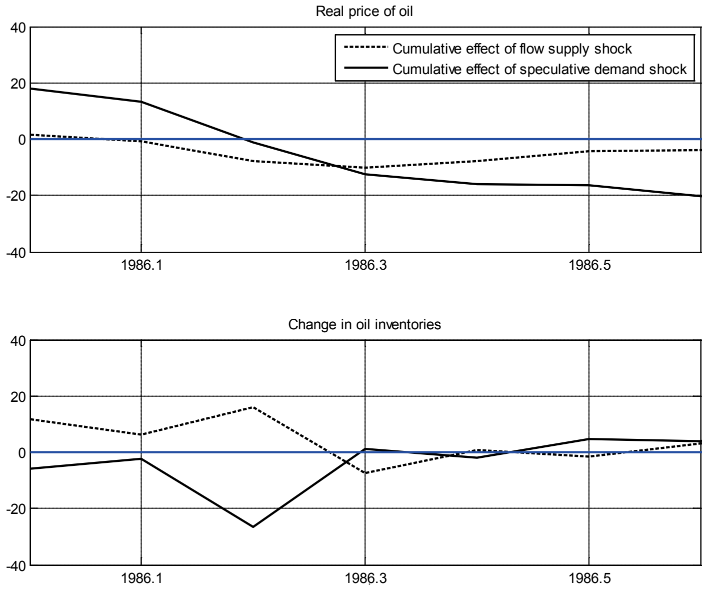
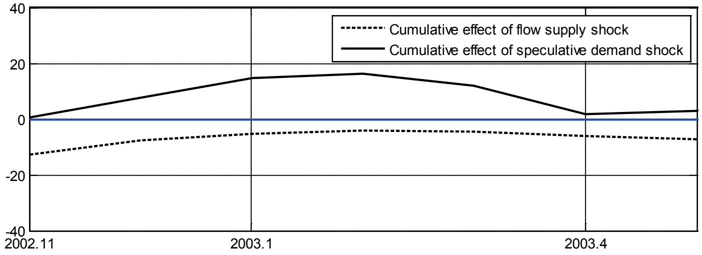
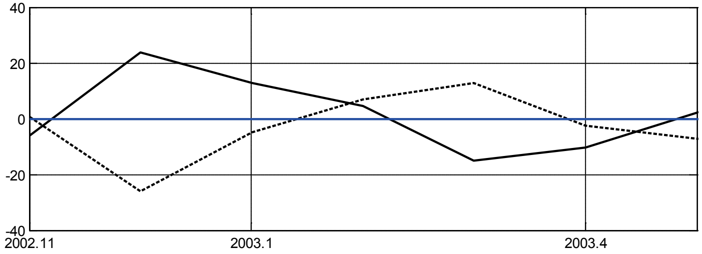


Figure III.7: Historical Decompositions for Venezuelan Crisis and Iraq War in 2002/03



Real price of oil



Change in oil inventories

Appendix III.A: Construction of the Price Elasticity of Oil Demand in Use

In this appendix we show how the oil demand elasticity in use can be approximated with the help of our structural model of the oil market. The amount of oil used in period t, denoted by U_t, equals the quantity of oil produced in that period (Q_t) minus the oil that is added to the stock of inventories (ΔS_t):

$$U_t = Q_t - \Delta S_t.$$

The change in oil used over time therefore equals the change in oil produced minus the change in the addition to inventory stocks: $\Delta U_t = \Delta Q_t - \Delta^2 S_t$. The price elasticity of oil demand in use is defined as:

$$\eta_t^{Use} \equiv \frac{\%\Delta U_t}{\%\Delta P_t} = \frac{\dfrac{\Delta Q_t - \Delta^2 S_t}{Q_{t-1} - \Delta S_{t-1}}}{\%\Delta P_t},$$

where Δ represents changes, $\%\Delta$ indicates percent changes in response to an oil supply shock in period t, and P_t denotes the real price of oil. Denote by $\tilde{B}_{11}$ the impact response of the percent change in oil production to an oil supply shock, where $\tilde{B}_{ij}$ refers to the ij th element of $\tilde{B}$. Then the implied change in oil production is $\Delta Q_t = Q_{t-1} \times (1 + \tilde{B}_{11}/100) - Q_{t-1} = Q_{t-1} \times \tilde{B}_{11}/100.$ Moreover, $\Delta^2 S_t = \Delta S_t - \Delta S_{t-1} = \overline{\Delta S} + \tilde{B}_{41} - \overline{\Delta S} = \tilde{B}_{41}$, where the change in oil inventories in response to the oil supply shock equals the impact response $\tilde{B}_{41}$ and, prior to the shock, the change in crude oil inventories is equal to its mean $\overline{\Delta S}$, which is observable. Finally, the impact percent change in the real price of oil in response to an oil supply shock is $\tilde{B}_{31}$. Hence, the demand elasticity in use can be expressed equivalently as

$$\eta_t^{Use} = \frac{\dfrac{(Q_{t-1} \times \tilde{B}_{11}/100) - \tilde{B}_{41}}{Q_{t-1} - \overline{\Delta S}}}{\tilde{B}_{31}/100}.$$

Note that by construction η_t^{Use} depends on Q_{t-1} and hence will be time-varying even though the oil demand elasticity in production is not. We therefore report the average oil demand elasticity in use over the sample period throughout this paper, denoted by $\eta^{O,Use}$.

Appendix III.B: Linking the Short-Run Price Elasticities of Gasoline and Oil Demand

In this appendix we derive an explicit relationship between consumers' demand for gasoline and refiners' demand for crude oil in a model in which refiners are allowed to, but not required to have market power in the gasoline market. Refiners are treated as price-takers in the crude oil market. Our analysis is strictly short-term, as is appropriate in constructing impact price elasticities. In the interest of tractability, we abstract from the fact that gasoline is only one of several refined products jointly produced from crude oil. We postulate that gasoline is produced according to a Leontief production function over capital, labor, and oil, $G = \min(K, L, \alpha O)$. If capital is fixed in the short run and refiners' labor input can be varied on the intensive margin, which seems plausible in practice, refiners produce gasoline in fixed proportion to the quantity of oil consumed, $G = \alpha O$, and pay a marginal cost equal to the price of oil, P_O, plus the marginal cost of labor, $MC = P_O + c$. P_G denotes the price of gasoline.

Consumers demand $G(P_G) = XP_G^{-\sigma} / P^{1-\sigma}$, where X is the expenditure on gasoline, P is the consumer price index, and the price elasticity of demand for gasoline, η^G, equals $-\sigma$. The inverse demand function is $P_G(G) = \omega G^{-1/\sigma}$ where $\omega \equiv X / P^{1-\sigma}$. In the Cournot-Nash equilibrium, each of J identical refinery firms will choose its own quantity of gasoline output, g_j, $j = 1, ..., J$, given the outputs of other firms, to maximize profits $\pi_j = P_G(G) g_j - (c + P_O) g_j$ with respect to g_j, where $G = \sum_j g_j$. The first-order condition is $\omega G^{-1/\sigma} - \omega g_j G^{(-\sigma-1)/\sigma} / \sigma - (c + P_O) = 0$, $j = 1, ..., J$. Summing over j and solving for the market price and gasoline production yields

$$P_G = \frac{J\sigma(P_O + c)}{J\sigma - 1} \qquad G = \left(\frac{\omega(J\sigma - 1)}{J\sigma(P_O + c)} \right)^{\sigma}.$$

Given $G = \alpha O$, we obtain

$$\alpha O = \left(\frac{\omega(J\sigma - 1)}{J\sigma(P_O + c)} \right)^{\sigma}.$$

Log-linearization yields $\eta^{O,Use} \approx \eta^G P_O / (P_O + c)$, where $\eta^{O,Use}$ denotes the price elasticity of demand for crude oil in use. The marginal cost estimates in Considine (1997) suggest that $c \approx 0$, which implies $\eta^{O,Use} \approx \eta^G$.

References

Alquist, R., and L. Kilian (2010), "What Do We Learn from the Price of Crude Oil Futures?" *Journal of Applied Econometrics*, 25, 539-573.

Barsky, R.B., and L. Kilian (2002), "Do We Really Know that Oil Caused the Great Stagflation? A Monetary Alternative," in: *NBER Macroeconomics Annual 2001*, B.S. Bernanke and K. Rogoff (eds.), MIT Press: Cambridge, MA, 137-183.

Baumeister, C., and L. Kilian (2012a), "Real-Time Forecasts of the Real Price of Oil," *Journal of Business and Economic Statistics*, 30, 326-336.

Baumeister, C., and L. Kilian (2012b), "Real-Time Analysis of Oil Price Risks Using Forecast Scenarios," mimeo, University of Michigan.

Baumeister, C., and G. Peersman (2012), "The Role of Time-Varying Price Elasticities in Accounting for Volatility Changes in the Crude Oil Market," forthcoming: *Journal of Applied Econometrics*.

Bodenstein, M., and L. Guerrieri (2011), "On the Sources of Oil Price Fluctuations," mimeo, Federal Reserve Board.

Burger, N.E., and D.T. Kaffine (2009), "Gas Prices, Traffic, and Freeway Speeds in Los Angeles," *Review of Economics and Statistics*, 91, 652-657.

Canova, F. and D. De Nicolo (2002), "Monetary Disturbances Matter for Business Cycle Fluctuations in the G-7," *Journal of Monetary Economics*, 49, 1131-1159.

Caballero, R.J., E. Farhi, and P.-O. Gourinchas (2008), "Financial Crash, Commodity Prices, and Global Imbalances," *Brookings Papers on Economic Activity*, Fall, 1-55.

Considine, T.J. (1997), "Inventories under Joint Production: An Empirical Analysis of Petroleum Refining," *Review of Economics and Statistics*, 79, 493-502.

Cooper, J.C.B. (2003), "Price Elasticity of Demand for Crude Oil: Estimates for 23 Countries," *OPEC Review*, 27, 1-8.

Dahl, C.A. (1993), "A Survey of Oil Demand Elasticities for Developing Countries," *OPEC Review*, 17, 399-419.

Dahl, C.A., and T. Sterner (1991), "Analyzing Gasoline Demand Elasticities: A Survey," *Energy Economics*, 13, 203-210.

Davis, L.W., and L. Kilian (2011), "Estimating the Effect of a Gasoline Tax on Carbon Emissions," forthcoming: *Journal of Applied Econometrics.*

Dvir, E., and K. Rogoff (2010), "Three Epochs of Oil," mimeo, Harvard University.

Edelstein, P., and L. Kilian (2009), "How Sensitive Are Consumer Expenditures to Retail Energy Prices?" *Journal of Monetary Economics*, 56, 766-779.

Fattouh, B., L. Kilian and L. Mahadeva (2012), "The Role of Speculation in Oil Markets: What Have We Learned So Far?" mimeo, University of Michigan.

Fry, R. and A.R. Pagan (2011), "Sign Restrictions in Structural Vector Autoregression: A Critical Review," *Journal of Economic Literature*, 49, 938-960.

Gause, G.F. (2002), "Iraq's Decision to Go to War, 1980 and 1990," *Middle East Journal*, 56, 47-71.

Giannone, D., and L. Reichlin (2006), "Does Information Help Recover Structural Shocks from Past Observations," *Journal of the European Economic Association*, 4, 455-465.

Hamilton, J.D. (2009a), "Causes and Consequences of the Oil Shock of 2007-08," *Brookings Papers on Economic Activity*, 1 (Spring), 215-261.

Hamilton, J.D. (2009b), "Understanding Crude Oil Prices", *Energy Journal*, 30, 179-206.

Hamilton, J.D., and A.M. Herrera (2004), "Oil Shocks and Aggregate Economic Behavior: The

Role of Monetary Policy," *Journal of Money, Credit and Banking*, 36, 265-286.

Hausman, J.A., and W.K. Newey (1995), "Nonparametric Estimation of Exact Consumers' Surplus and Deadweight Loss," *Econometrica*, 63, 1445-1476.

Hughes, J.E., Knittel, C.R., and D. Sperling (2008), "Evidence of a Shift in the Short-Run Price Elasticity of Gasoline Demand," *Energy Journal*, 29, 113-134.

Inoue, A., and L. Kilian (2011), "Inference on Impulse Response Functions in Structural VAR Models," mimeo, University of Michigan.

Kellogg, R. (2011), "Short-Run Responses of Oil Production to Spot and Future Oil Prices," mimeo, University of Michigan.

Kilian, L. (2008a), "Exogenous Oil Supply Shocks: How Big Are They and How Much Do They Matter for the U.S. Economy?" *Review of Economics and Statistics*, 90, 216-240.

Kilian, L. (2008b), "The Economic Effects of Energy Price Shocks," *Journal of Economic Literature*, 46, 871-909.

Kilian, L. (2009a), "Not all Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market," *American Economic Review*, 99, 1053-1069.

Kilian, L. (2009b), "Comment on 'Causes and Consequences of the Oil Shock of 2007-08' by James D. Hamilton," *Brookings Papers on Economic Activity*, 1 (Spring), 267-278.

Kilian, L., and B. Hicks (2012), "Did Unexpectedly Strong Economic Growth Cause the Oil Price Shock of 2003-2008?" forthcoming: *Journal of Forecasting*.

Kilian, L., and D. Murphy (2012), "Why Agnostic Sign Restrictions Are Not Enough: Understanding the Dynamics of Oil Market VAR Models," forthcoming: *Journal of the European Economic Association*.

Rubio-Ramirez, J., Waggoner, D., and T. Zha. (2010) "Markov-Switching Structural Vector Autoregressions: Theory and Application", *Review of Economic* Studies, 77, 665-696.

Serletis, A., G. Timilsina, and O. Vasetsky (2010), "International Evidence on Sectoral Interfuel Substitution," *Energy Journal*, 31, 1-29.

Sockin, M., and W. Xiong (2012), "Feedback Effects of Commodity Futures Prices," mimeo, Princeton University.

Sweeney, J.L. (1984), "The Response of Energy Demand to Higher Prices: What Have We Learned?" *American Economic Review*, 74, 31-37.

Terzian, P. (1985), *OPEC. The Inside Story*. Zed Books: London.

Uhlig, H. (2005) "What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics*, 52, 381-419.

Yatchew, A., and J.A. No (2001), "Household Gasoline Demand in Canada," *Econometrica*, 69, 1697-1709.

Yergin, D. (1992), *The Prize. The Epic Quest for Oil, Money, and Power.* New York: Simon and Schuster.