THREE ESSAYS ON INTERNATIONAL ECONOMICS

by

SHU-WING EDDERY LAM

B.A., Boston University, 2003
M.A., Boston University, 2003
M.A., University of Massachusetts at Amherst, 2006

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Economics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2012

Abstract

This dissertation comprises of three essays in international macroeconomics. The first essay investigates the competition between two city states, both of which will stand in place of countries in the global scheme. Under the framework of the three-stages-game, we assume that there are two cities competing for dominance over two sectors: the manufacturing sector and the financial sector. In addition, the government of each city state can build infrastructure to increase the competitiveness of the financial and distributive firms of its city. Under this framework, we are able to show that the amount of resources, the start-up costs of providing services, and the relative effectiveness of their infrastructures determine the optimal amounts of infrastructures the cities decide to build, and thus also decide the equilibrium outcome of this game.

In my second essay, we examine the relationship between income distribution and import patterns. The Linder hypothesis states that countries with similar economic characteristics should trade more often. However, although the total volumes of trade between these countries are similar, the traded goods may be different. This paper investigates the trading patterns of countries with similar characteristics. Specifically, we analyze the relationship between the import patterns and income distributions of importers. We develop an import similarity index to portray the composition of imports and utilize the idea of a "market overlap," a theoretical concept proposed by Bohman and Nilsson (2007), to represent the similarity of income distributions across different importing countries. We provide empirical evidence to support the notion that countries with similar income distributions display similar import patterns. We also separate countries by income level and find that income distribution exerts a positive impact on the similarity of import patterns for all but low income countries. Finally, we incorporate the characteristics of goods into our analysis and

show that the positive relationship between income distributions and import patterns holds for differentiated and reference-priced goods, but not for homogeneous goods.

In my final essay, we look into another aspect of international literature: the exchange rate. In the literature, we find that vector autoregressive (VAR) models and impulse response analyses are common tools to study the relationship between monetary policy and exchange rate movements. Therefore, it is important to investigate the accuracy of the VAR model. In the first part of this essay, we assume that the true, underlying, data-generating process is hump-shaped, which is the shape of the impulse response of exchange rate to a monetary policy shock. We show that results estimated from any VAR models applying AIC as their lags selection are biased. We also introduce two possible solutions to remedy this bias: the use of more lags in the VAR models or the use of the proposed loss functions estimations. These results suggest we should be cautious when interpreting empirical evidences on international literature.

In the second part of the same essay, we investigate another issue that is closely related to the exchange rate and the VAR model. Under the estimation of the VAR model, the researcher implicitly assumes that the objective loss function is quadratic. However, it is a well accepted fact that monetary authority adjusts the interest rate according to policy. One of the objectives of the monetary authority is to influence the exchange rate in their favor. They estimate the size of the loss caused by deviations from the current exchange rate to the rate they desire, and then they adjust the amount of money in the international market. We propose an asymmetric loss function that monetary authorities may use to estimate the impulse response of the exchange rate to a contractionary monetary policy shock. We then compare these estimated impulse response functions to those estimated by the VAR. We find that while both of these estimated impulse response functions share the same sign, the magnitude and the duration of the shock are quite different. These results suggest that the VAR model may not be appropriate in estimating the exchange rate movement.

THREE ESSAYS IN INTERNATIONAL ECONOMICS

by

SHU-WING EDDERY LAM

B.A., Boston University, 2003
M.A., Boston University, 2003
M.A., University of Massachusetts at Amherst, 2006

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Economics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2012

Approved by:

Co-Major Professor
Dr. Lance Bachmeier

Approved by:

Co-Major Professor
Dr. Yang-Ming Chang

Copyright

Eddery Lam

2011

Abstract

This dissertation comprises of three essays in international macroeconomics. The first essay investigates the competition between two city states, both of which will stand in place of countries in the global scheme. Under the framework of the three-stages-game, we assume that there are two cities competing for dominance over two sectors: the manufacturing sector and the financial sector. In addition, the government of each city state can build infrastructure to increase the competitiveness of the financial and distributive firms of its city. Under this framework, we are able to show that the amount of resources, the start-up costs of providing services, and the relative effectiveness of their infrastructures determine the optimal amounts of infrastructures the cities decide to build, and thus also decide the equilibrium outcome of this game.

In my second essay, we examine the relationship between income distribution and import patterns. The Linder hypothesis states that countries with similar economic characteristics should trade more often. However, although the total volumes of trade between these countries are similar, the traded goods may be different. This paper investigates the trading patterns of countries with similar characteristics. Specifically, we analyze the relationship between the import patterns and income distributions of importers. We develop an import similarity index to portray the composition of imports and utilize the idea of a "market overlap," a theoretical concept proposed by Bohman and Nilsson (2007), to represent the similarity of income distributions across different importing countries. We provide empirical evidence to support the notion that countries with similar income distributions display similar import patterns. We also separate countries by income level and find that income distribution exerts a positive impact on the similarity of import patterns for all but low income countries. Finally, we incorporate the characteristics of goods into our analysis and

show that the positive relationship between income distributions and import patterns holds for differentiated and reference-priced goods, but not for homogeneous goods.

In my final essay, we look into another aspect of international literature: the exchange rate. In the literature, we find that vector autoregressive (VAR) models and impulse response analyses are common tools to study the relationship between monetary policy and exchange rate movements. Therefore, it is important to investigate the accuracy of the VAR model. In the first part of this essay, we assume that the true, underlying, data-generating process is hump-shaped, which is the shape of the impulse response of exchange rate to a monetary policy shock. We show that results estimated from any VAR models applying AIC as their lags selection are biased. We also introduce two possible solutions to remedy this bias: the use of more lags in the VAR models or the use of the proposed loss functions estimations. These results suggest we should be cautious when interpreting empirical evidences on international literature.

In the second part of the same essay, we investigate another issue that is closely related to the exchange rate and the VAR model. Under the estimation of the VAR model, the researcher implicitly assumes that the objective loss function is quadratic. However, it is a well accepted fact that monetary authority adjusts the interest rate according to policy. One of the objectives of the monetary authority is to influence the exchange rate in their favor. They estimate the size of the loss caused by deviations from the current exchange rate to the rate they desire, and then they adjust the amount of money in the international market. We propose an asymmetric loss function that monetary authorities may use to estimate the impulse response of the exchange rate to a contractionary monetary policy shock. We then compare these estimated impulse response functions to those estimated by the VAR. We find that while both of these estimated impulse response functions share the same sign, the magnitude and the duration of the shock are quite different. These results suggest that the VAR model may not be appropriate in estimating the exchange rate movement.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I am grateful for my committee members who have contributed greatly to the learning process during my PhD program and the completion of my dissertation. To the co-chairs of the committee, Dr. Lance Bachmeier and Dr. Yang-Ming Chang, I am thankful for your infinite patience and support in both the programming and the professional editing of this dissertation. The rest of the committee has also been spectacularly supportive during this process. Dr. Charlie Braymen worked closely with me in the international trade area and was always available to meet with me and discuss my dissertation despite his busy schedules. Dr. Neill and Dr. Tian offered invaluable comments and advice toward the completion of my dissertation. To all the professors in the economics department who have taught and guided me in countless ways, I would like to say, thank you.

I would also like to recognize the great friends I have made during my graduate studies: Shane Sanders, Bhavneet Walia, Ruben Sargsyan, Canh Le, Bandar Aba-Alkhail, Renfeng Xiao, Yuan Gao, Casey Abington, Yaseen Alhaj, Jaime Andersen, Dave Brown, Kyle Ross, Abhinav Alakshendra, Yacob Zereyesus, Kara Ross, Bebonchu Atems, Vladimir Bejan, Laura Chen, Zijun Luo, Mercy Palamuleni, Patrick Scott, Jenny Wu, andChris Youderian, Koichi Yamaura as well as many others whom I have not mentioned.

I would also like to thanks everyone in the economics department and the many friends who have come into my life at Kansas State University and in Manhattan whom I have failed to mention, but contributed in any form towards the successful completion of this dissertation.

Finally, I am grateful for my family for supporting me throughout my life and believing in me. To my parents who have always had a great faith in me, thanks for standing beside me. To my siblings Francis and Jessica, thank you for your moral support and encouragement throughout all of these years.

Dedication

I dedicate this dissertation to my family.

Chapter 1

On the Economics of Inter-City Competition In Financial and Distribution Markets

1.1   Introduction.

Cities constitute an important economic jurisdiction in each country due to the fact that city or local governments provide basic infrastructures for facilitating business activities. Cities are the most populous and the most prosper areas in a country. It is arguably that without the development of cities, civilizations would not have places to nurture and the world economy may not be as prosper as they are today. Interestingly, cities compete on various economic dimensions (to be a financial center, a distribution headquarter, an Olympic City, etc.), whether the cities are located within a country or in different nations. The issues of inter-city competition across national boundaries have increasingly becoming more important in recent decades.

There are voluminous studies on the economics of cities. In general, these literatures can be divided into two categories: cities can either compete against each other, or they can be cooperative and mutually benefit each other. In the case of competition, the cities will want to undercut its rivalries but lowering their production costs. For example, a recent contribution by Long and Wong (2009) shows that governments provide infrastructures which lower the costs for local firms for the purpose of enhancing their competitiveness. In

the case of cooperation, the cities can trade inputs, intermediate goods and final goods to each other. Thus their main concerns are rather cities should diversify their productions and how many labor should be in each cities. Anas and Xiong (2003) shows that the transportation costs of these inputs and the final goods determine whether cities should diversify their productions. If the transportation costs of inputs are low, cities should diversify their production. On the other hand if the transportation costs of final goods are low, cities should specialize their productions. In Cavailhesa et al. (2007) they show that trade cost, communication cost and commuting cost determine the size of the cities and how firms allocate within cities. Several industrial districts may arise within the same city in their model. Henderson (1974) develops a theoretical model to analyze the optimal size of cities. In his paper the role of capital owners determine the optimal size of cities. If capital owners are also in labor force, then the optimal city size tends to be smaller. On the other hand, if the capital owners are purely investors, conflict of interests between labor forces and investors may lead to undesirable equilibrium, that is, equilibrium city size may not be identical to optimal city size. Finally, Grajeda and Sheldon (2009) provide empirical study on the relationship between trade openness and the size of the city. Their results are not consistent, depending on the econometric models they estimated and the data they used. It appears that relatively little research in the economics literature has been conducted to analyze the economic geography of inter-city competition between nations. The recent contribution by Long and Wong (2009) is an exception. They develop a spatial model to analyze strategic rivalry between cities located in different countries to become the major distribution center in a region.

This paper extends the single-dimension framework of city competition as developed by Long and Wong (2009). In our analysis, we explicitly consider resources constraints facing two city governments in a setting that involves multiple dimensions. Specifically, each government maximizes its objective function by strategically distributing limited resources into two different sectors: financial sector and production sector. Governments can assist

2

their firms in these sectors by spending resources to build respective infrastructures. These infrastructures, in return, reduce the service costs of respective sectors and increase the competitiveness of firms in the sectors. We wish to analyze city competition in two different sectors under sequential game structure. While it would be desirable to use the simultaneous game structure as they did in their paper, the complexity and intractability of mathematics prevents us to do so and therefore, we decide to solve this model under sequential game structure. We show that if the resources of one city is relatively abundant than the other city, ceteris paribus, this city tends to dominate both the financial and distribution sectors. Given that the amount of resources a city has may reflect the size and the prosperity of the city, our analysis implies that larger and wealthier cities are more likely to become both the distribution and financial centers. We further derive conditions under which one city dominates the financial sector whereas the other city dominates the distribution sector.

The remainder of this paper is organized as follows. Section 2 presents the analytical framework of competition between two cities in financial and distribution markets. In Section 3, we examine the case in which one city captures both of the two markets. Sections 4 and 5 analyze duopolistic competition in the financial market and in the distribution market, respectively. In Section 6, we discuss how government resource constraints affect the dominance of their firms in the financial and distribution markets. Concluding remarks can be found in Section 7.

1.2 The Model

Following Long and Wong (2009), we analyze competition between two cities, denoted as city N (North) and city S (South), in which the distance between them is normalized at unity. City S is located at point 0 and city N is located at point 1. There exists a continuum of investors/producers who are uniformly distributed between the two cities. Each city is assumed to have a single distribution firm. To allow for inter-city competition in multiple dimensions, we further incorporate a financial sector into our model by considering that

each investor has to acquire a loan from a financial firm (located in city S or N) to finance production. For analytical simplicity, there is a single financial firm in each city. Assuming that each investor produces one unit of a homogenous good, the investor has to decide whether to sell the product locally or to export it. If an investor decides to export the good, he has to hire a distribution firm (referred to as distributor S or N) to distribute it to the rest of the world. The world price of this homogenous good is taken to be higher than its local price.

In offering a production loan to each investor, financial firm S or N charges α_i, and incurs a service cost C_i^F ($<\alpha_i$), where $i = $ S or N. It is costly for an investor to collect information about financial services and this cost is positively related to where the investor located. For instance, the investors have to travel to a city for negotiation or signing contracts. If an investor is located at point x, the cost of travelling to city S is $\frac{C}{2}x^2$. This cost can be considered as acquiring or researching information about the financial firm in city S. This cost is $\frac{c}{2}(1-x)^2$ if the investor x obtains information from financial firm N.

In providing service of distributing the good for to an investor, distributor S or N charges θ_i and incurs a cost C_i^D ($<\theta_i$), where $i = $ S or N. If an investor is located at point y, the cost of shipment to city S is $\frac{b}{2}y^2$ and that of shipment to city N is $\frac{b}{2}(1-y)^2$.

Finally, there is a government in each of the two cities. A city government (referred to as government S or N) can help its financial and distribution firms to become more competitive by building infrastructures. By infrastructures, we do not limit ourselves to physical infrastructure such as new airport. Sophisticated legal and financial systems certainly can help reducing the operation costs of a financial firm and improvements on them are also considered as infrastructures. We consider two types of infrastructures: I_i^D and I_i^F. Infrastructure I_i^F helps to reduce the financial firm's service cost, and infrastructure I_i^D helps to cut the distribution firm's service cost. Specifically, we assume that

$$C_i^F = \overline{C_i^F} - \frac{\psi_i^F (I_i^F)^2}{2} \tag{1.1}$$

4

$$C_i^D = \overline{C_i^D} - \frac{\psi_i^D (I_i^D)^2}{2} \tag{1.2}$$

where C_i^j is the service cost of firm j (= F, D) in city i(= N, S), and $\bar{C_i}^j$ is firm $j's$ service cost in the absence of infrastructures built by the governments. Symbol ψ_i^j reflects the effectiveness of infrastructure I_i^j in reducing costs of providing services by firm j in city i . The specifications in equations (1.1) and (1.2) indicate that each city government is capable of lowering operating costs for their firms by building more efficient infrastructures. Given the limited amount of resources available, however, each government has to decide by how much and which particular type of infrastructures to build.

1.2.1   The Three-Stage Game

We consider a three-stage sequential-move game. We will offer more detail explanations and objectives in each stage of the game. Note that if an investor decides to export his product to the rest of the world, he has to hire both financial and distribution firms. However, if an investor finds out that the costs of hiring these firms outweigh the benefits, he decides not to undertake the investment. It is possible for an investor to hire a financial firm only. In this case the investor decides to sell his unit of good locally.

First Stage

At the beginning of this game each city government determines the types and amounts of infrastructures to build. We assume that the main objective of government S or N is to maximize the total profit of its financial and distribution firms. As mentioned above, a city government can help both types of its firms to become more competitive by building respective infrastructures. These infrastructures help reduce the service costs of respective firms, allowing them to offer lower prices to the investors. However, each government has a limited amount of resources. Denote R_i as the (exogenous) amount of resources available to government i (= S, N). The objective of government i is to choose I_i^F and I_i^D that maximize

the following constrained optimization problem:

$$\max_{\{I_i^F, I_i^D\}} \left(\pi_i^F + \pi_i^D \right) \qquad (1.3)$$

subject to

$$R_i \geq I_i^F + I_i^D \qquad (1.4)$$

where π_i^j is the profit of firm j in city j. Resources R_i can be used to reflect city size. If the sizes of both cities and the effectiveness of their infrastructures are similar, it is likely that cities S and N share equally both the financial and distribution markets. However, if one city is much bigger than the other (thus more resources), it is likely that the big city captures both markets. In the later section we will derive conditions that determine which scenarios will emerge.

Second Stage

In the second stage of the game, each financial firm decides how many investors to serve, as well as an optimal price to charge. Recall that an investor has to acquire a loan from a financial firm to finance production. If an investor finds that production is unprofitable, he will not apply for any loan. This condition limits the maximum price the financial firm can charge to the investors and hence constitutes a constraint for the firm's profit maximization decision. Denote α_s as the price financial firm S charges to the investors and x_S as the number of investors served by the firm. Let P be the competitive price the investors can sell their goods locally. The objective of financial firm S is to choose α_s and xs to solve the following problem:

$$\max_{\{\alpha_S, x_S\}} \pi_S^F = \int_0^{x_s} \left(\alpha_S - C_S^F \right) dx = \left(\alpha_S - C_S^F \right) x_S \qquad (1.5)$$

subject to

$$P \geq \alpha_S + \frac{C}{2}x_S^2 \tag{1.6}$$

For analytical simplicity, we assume that the financial firm's service cost, C_S^F, is independent to the number of investors served by the firm. But this service cost is a function of infrastructure $I_S{}^F$ as described by equation (1.1). Note that $\frac{C}{2}x_S^2$ in equation (1.6) is the cost to the investor at location x_S of obtaining information about financial firm S.

Similarly, we can rewrite the profit maximization problem that financial firm N faces by replacing x_S with (1- x_S) and by changing the subscript to N.

Third Stage

In the third and last stage of the game, distributor S or N has to decide how many investors to serve, at what prices. In our setting, the number of investors to be served by each distributor depends on the equilibrium outcome of the second stage. If an investor decides not obtain any loan from a financial firm, he will not hire any distributor either simply because he is not producing.

Let $P*$ be the competitive price of the homogeneous good in the world market. Under the assumption that $P* > P$, there is an incentive for the investors to export. Given the service prices charged by the distributors, if the investors realize that export is unprofitable, they will not hire the distribution firms. This condition imposes a constraint on the price that a distributor can charge. The constrained profit maximization problem that distributor S solves is:

$$\max_{\{\theta_S, y_S\}} \pi_S^D = \int_0^{y_S} \left(\theta_S - C_S^D\right) dy = \left(\theta_S - C_S^D\right) y_S \tag{1.7}$$

subject to

$$P* - \left(\alpha_S + \frac{C}{2}x_S^2\right) \geq \theta_S + \frac{b}{2}y_S^2 \tag{1.8}$$

Similarly, we can set up the constrained profit maximization problem for distributor N by replacing y_S with $(1 - y_s)$ and by changing the subscript to N.

Note that y_S is the number of investors served by distributor S. We also assume that the a distributor's service cost, C_S^D, is independent of y_S. But this service cost is a function of infrastructure I_S^D as given by equation (1.2). The term θ_S is the price that distributor S charges. The term $\frac{b}{2}y_S^2$ is the cost of shipping the good to city S. The budget constraint in (1.8) is slightly different from that for a financial firm as discussed above. Since an investor has to acquire a loan from a financial firm for production, the investor has to subtract the cost of hiring a financial firm, $\left(\alpha_i + \frac{c}{2}x_i^2\right)$, in calculating profit from export. The budget constraint holds only when an investor hires distributor S or otherwise the investor will just sell his product locally.

1.3 Monopolistic Financial Firm and Distribution Firm

As in game theory, we use backward induction to solve the three-stage game. We focus our analysis temporarily on a single city case by assuming that the financial and distribution firms in city N are inefficient to the point where they do not offer any services. In this case, financial firm S and distributor S become monopolistic service providers. Three possibilities of interest to our analysis are as follows:

1. Both financial firm S and distributor S serve their respective markets completely;

2. Financial firm S serves its market completely, but distributor S serves its market partially;

3. Both financial firm S and distributor S serve their respective markets partially[1].

---

[1]Notice that under this setting it is not possible to have a case where the distribution firm serves the whole market and financial firm serves only the partial market. Investors are required to hire a financial firm if they want to start producing, but since the financial firm decides only to serve partial market, some investors must have left out and they will not hire distribution firm since they are not producing at all.

In what follows, we discuss the whole market case and the partial market case separately. We begin our analysis with distributor S at stage three and then financial firm S at stage two.

Third Stage: Distributor S

Distributor S is able to serve the entire distribution market if and only if:

$$\frac{3b}{2} \leq P* - C_S^D - \frac{C}{2}x_S^2 - \alpha_S \tag{1.9}$$

This is the scenario where $y_s = 1$. Given this condition, we derive the price that distributor S charges to the investor by solving the constrained profit optimization problem (see equations (1.7) and (1.8)):

$$\theta_S = P* - \frac{C}{2}x_S^2 - \alpha_S - \frac{b}{2} \tag{1.10}$$

Distributor S serves the market only partially if and only if:

$$\frac{3b}{2} \geq P* - C_S^D - \frac{C}{2}x_S^2 - \alpha_S \tag{1.11}$$

In this scenario, Distributor S determines θ_s and y_s that maximize profit in equation (1.7) subject to the constraint in equation (1.8). The solutions are:

$$\theta_S = \frac{2}{3}\left[P* - \alpha_S\right] + \frac{1}{3}\left[C_S^D - Cx_S^2\right] \tag{1.12}$$

$$y_S = \sqrt{\frac{1}{b}\left[\frac{2}{3}(P* - \alpha_S - C_S^D) - \frac{1}{3}x_S^2\right]} \tag{1.13}$$

In the above two cases, service price θ_s depends on x_s, the number of investors served by distributor S. The distributor has to take into account the service price that financial firm S charges, α_S, in the second stage of the three-stage game.

Second Stage: Distributor S

Financial firm S is able to serve the entire financial market if and only if:

$$\frac{3C}{2} \leq P - C_S^F \tag{1.14}$$

This is the scenario in which $x_S = 1$. Given this condition, we solve the constrained profit maximization problem (see equations (1.5) and (1.6)) for the price that financial firm S charges to the investors:

$$\alpha_S = P - \frac{C}{2} \tag{1.15}$$

Financial firm S serves the market only partially if and only if:

$$\frac{3C}{2} \geq P - C_S^F \tag{1.16}$$

Financial firm S determines α_S and x_S that maximize profit in equation (1.5) subject to the constraint in equation (1.6). The solutions are:

$$\alpha_S = \frac{2}{3}P + \frac{C_S^F}{3} \tag{1.17}$$

$$x_S = \sqrt{\frac{2}{3C}(P - C_S^F)} \tag{1.18}$$

Case I: Both Firms in City S Serve Their Respective Markets Completely

In this case, $x_S = 1$ and $y_S = 1$. The service price charged by financial firm S is given in equation (1.15). To determine the service price θ_S charged by distributor S, we substitute equation (1.15) into equation (1.10) and set $x_S = 1$ to obtain

$$\theta_S = P* - P - \frac{b}{2}$$

10

We calculate profits for financial firm S and distributor S which are given, respectively, as:

$$\pi_S^F = P - \frac{C}{2} - C_S^F, \pi_S^D = P* -P - \frac{b}{2} - C_S^D \tag{1.19}$$

Case II: Only Financial Firms in City S Serve Its Markets Completely

In this case $x_S = 1$ and equation (1.15) also defines α_S. To obtain θ_S, we substitute equation (1.15) into equation (1.12) and set $x_S = 1$ to obtain

$$\theta_S = \frac{2}{3}[P* -P] + \frac{1}{3}C_S^D \tag{1.20}$$

To obtain y_S, we first set x_S to be 1 in equation (1.13). We then substitute equation (1.20) into equation (1.13) to get

$$y_S = \sqrt{\frac{1}{b}[\frac{2}{3}(P* -P) + \frac{C}{3} - \frac{1}{3}]} \tag{1.21}$$

We calculate profit for distributor S as follows:

$$\pi_S^D = \frac{2}{3}(P* -P - C_S^D)\sqrt{\frac{1}{b}[\frac{2}{3}(P* -P) + \frac{C}{3} - \frac{1}{3}]} \tag{1.22}$$

In this case, profit for financial firm S is identical to that in the previous case:

$$\pi_S^F = P - \frac{C}{2} - C_S^F \tag{1.23}$$

Case III: Both Firms in City S Serve Their Respective Markets Partially

In this case, α_S is given by the one in equation (1.17) and x_S is given by the one in equation (1.18). Profit for financial firm S is:

$$\pi_S^F = \frac{2}{3}(P - C_S^F)\sqrt{\frac{2}{3C}(P - C_S^F)} \tag{1.24}$$

11

To derive y_S, we substitute equations (1.17) and (1.18) into equation (1.13) to get

$$y_S = \sqrt{\frac{1}{b}(\frac{2}{3}[P*-P-C_S^D] + \frac{4}{9}C_S^F}$$

Similarly, to obtain θ_S, we substitute equations (1.17) and (1.18) into equation (1.12) to get

$$\theta_S = \frac{2}{3}(P*-P) + \frac{4}{9}C_S^F + \frac{1}{3}C_S^D$$

We calculate profit for distributor S as follows:

$$\pi_S^D = [\frac{2}{3}(P*-P) + \frac{4}{9}C_S^F - \frac{2}{3}C_S^D]\sqrt{\frac{1}{b}(\frac{2}{3}[P*-P-C_S^D] + \frac{4}{9}C_S^F} \tag{1.25}$$

First Stage (of Case I)

Here we only consider Case I in which financial firm S and distributor S serve their respective markets completely. Recall that their profit functions are given, respectively as

$$\pi_S^F = P - \frac{C}{2} - C_S^F \tag{1.26}$$

$$\pi_S^D = P*-P - \frac{b}{2} - C_S^D \tag{1.27}$$

where C_S^D and C_S^F are costs of providing services by the firms as defined in equations (1.1) and (1.2). The objective of government S is to choose I_S^F and I_S^D that solve the following profit maximization problem:

$$\max_{\{I_S^F, I_S^D\}} (\pi_S^F + \pi_S^D)$$

subject to

$$R_S \geq I_S^F + I_S^D,$$

where π_S^F and π_S^D are given in equations (1.26) and (1.27). Solving the above problem yields

$$I_S^F = \frac{R_S \psi_S^D}{\psi_S^F + \psi_S^D} \tag{1.28}$$

$$I_S^D = \frac{R_S \psi_S^F}{\psi_S^F + \psi_S^D} \tag{1.29}$$

The equilibrium service costs for financial firm S and distributor S are given, respectively as:

$$C_S^F = \overline{C_S^F} - \frac{\psi_S^F}{2} \left(\frac{R_S \psi_S^D}{\psi_S^F + \psi_S^D} \right)^2$$

$$C_S^D = \overline{C_S^D} - \frac{\psi_S^D}{2} \left(\frac{R_S \psi_S^F}{\psi_S^F + \psi_S^D} \right)^2$$

We can use above equations to calculate α_S and θ_S.

1.4   Duopoly in the Financial Market

We proceed to examine the case in which the financial firms of the two cities compete in the financial market. We consider the whole market scenario where every investor is completely served by the same financial firm (either S or N). It is necessary to determine the critical investor who is indifferent between financial firm S and financial firm N. The investor, denoted as x_d, will be indifferent if:

$$\alpha_S + \frac{C}{2} x_d^2 = \alpha_N + \frac{C}{2} (1 - x_d)^2 \tag{1.30}$$

Solving for x_d yields

$$x_d = \frac{1}{C} (\alpha_N - \alpha_S) + \frac{1}{2} \tag{1.31}$$

This equation is similar to equation (19) for the case of distribution firms discussed in Long and Wong (2009). There are three possibilities:

1. Case I: If $\alpha_N - \alpha_S \geq \frac{C}{2}$, then $x_d = 1$ and financial firm S serves the entire financial market. Financial firm S harges the service price:

$$\alpha_S = \alpha_N - \frac{C}{2}$$

13

if $\alpha_N < P$, and

$$\alpha_S = P - \frac{C}{2}$$

if $\alpha_N > P$.

2. Case II: If $\alpha_S - \alpha_N \geq \frac{C}{2}$, then $x_d = 0$ and financial firm S serves the entire financial market. Financial firm S harges the service price:

$$\alpha_N = \alpha_S - \frac{C}{2}$$

if $\alpha_S < P$, and

$$\alpha_N = P - \frac{C}{2}$$

if $\alpha_S > P$.

3. Case III: If $|\alpha_N - \alpha_S| < \frac{C}{2}$, then financial firm S serves the investors located at $(0, x_d)$ and financial firm N serves those at $(x_d, 1)$.

As illustrated by Long and Wong, we derive the reaction functions of both financial firm and distributor by solving the following problem:

$$\max \pi_i = \left(\alpha_i - C_i^F\right) x_c \qquad subject\ to\ P - \alpha_i - \frac{C}{2} x_c \geq 0$$

The reaction functions of the respective firms are:

$$\alpha_S = R_S(\alpha_N) = \frac{C}{4} + \frac{1}{2}[C_S^F + \alpha_N], \alpha_N = R_N(\alpha_S) = \frac{C}{4} + \frac{1}{2}[C_N^F + \alpha_S]$$

Notice that $R_S(\alpha_N)$ intersects $\alpha_N - \frac{C}{2}$ when $\alpha_N = C_S^F + \frac{3}{2}C$. In other words, if $\alpha_N^* > C_S^F + \frac{3}{2}C$, financial firm S captures the entire market. Similar deviation shows that if $\alpha_S^* > C_N^F + \frac{3}{2}C$, financial firm N captures the entire financial market. According to the limit pricing proposition as shown in Long and Wong, if condition (1.9) holds and $P > \alpha_N > C_S^F + \frac{3}{2}C$, then financial firm S captures the entire market by charging a service price lower than the monopoly price.

With the above limit pricing proposition we derive the equilibrium condition for the second stage of this game. Similar to Long and Wong, both financial firms at this second

stage complete in a Bertrand fashion by choosing service prices simultaneously. We obtain the similar results to those in Long and Wong. Assuming the condition (1.9) holds:

1. If $C_N^F \geq \alpha_N = C_S^F + \frac{3}{2}C$, financial firm S captures the entire market. According to the limit pricing theorem, if $C_N^F < P$, financial firm S charges the service price:

$$\alpha_S = C_N^F - \frac{C}{2}$$

But if $C_N^F < P$, the financial firm charges the monopolistic service price:

$$\alpha_S = P - \frac{C}{2}$$

2. Similarly, if $C_S^F \geq \alpha_S = C_N^F + \frac{3}{2}C$, financial firm N captures the entire market. According to the limit pricing theorem, if $C_s^F < P$, financial firm N charges the service price:

$$\alpha_N = C_S^F - \frac{C}{2}$$

But if $C_S^F < P$, the financial firm charges the monopolistic service price:

$$\alpha_N = P - \frac{C}{2}$$

3. Finally, if $|C_N^F - C_S^F| < \frac{3}{2}C$, none of the two financial firms captures the entire market. It follows from the reaction functions that we have

$$\alpha_S = \frac{C}{2} + \frac{1}{3}(2C_S^F + C_N^F)$$

$$\alpha_N = \frac{C}{2} + \frac{1}{3}(2C_N^F + C_S^F)$$

1.5 Duopoly in the Distribution Market

In this section, we consider cases in which a financial firm, either S or N, captures the entire financial market. In other words, partial financial market is not considered here. Furthermore, here we only consider the whole market solution for distribution market.

Similar to the previous section, our characterize the investor who is indifferent between distributor S or distributor N. The investor, located at point y_d, will be indifferent if and only if:

$$\theta_S + \frac{b}{2}y_d^2 + \alpha_i + \frac{C}{2}x_i^2 = \theta_N + \frac{b}{2}(1 - y_d)^2 + \alpha_i + \frac{C}{2}x_i^2 \tag{1.32}$$

To investor y_d, the cost of hiring distributor S is identical to the cost of hiring distributor N. Note that on both sides of equation (1.32), the financial service price, α_i, and the cost of obtaining information, $\frac{C}{2}x_i^2$, are cancelled out. This is based on the assumption that at the second stage of the game there is only one financial firm (either S or N) capturing the entire financial market. Solving for y_d yields[2]

$$y_d = \frac{1}{b}(\theta_N - \theta_S) + \frac{1}{2}$$

Here there are three interesting possibilities, given that only one financial firm dominates in the financial market: 1. If $C_N^D \geq \alpha_N = C_S^D + \frac{3}{2}C$, financial firm S captures the entire market. According to the limit pricing theorem, if $C_N^D < P$, financial firm S charges the service price:

$$\theta_S = C_N^D - \frac{b}{2}$$

But if $C_N^D < P$, the financial firm charges the monopolistic service price:

$$\theta_S = P - \frac{b}{2}$$

---

[2]This result is similar to that is Long and Wong.

2. Similarly, if $C_S^D \geq \alpha_S = C_N^D + \frac{3}{2}C$, financial firm N captures the entire market. According to the limit pricing theorem, if $C_s^F < P$, financial firm N charges the service price:

$$\theta_N = C_S^D - \frac{b}{2}$$

But if $C_S^D < P$, the financial firm charges the monopolistic service price:

$$\theta_N = P - \frac{b}{2}$$

3. Finally, if $|C_N^D - C_S^D| < \frac{3}{2}C$, none of the two financial firms captures the entire market. It follows from the reaction functions that we have

$$\theta_S = \frac{b}{2} + \frac{1}{3}(2C_S^D + C_N^D)$$

$$\theta_N = \frac{b}{2} + \frac{1}{3}(2C_N^D + C_S^D)$$

where $P = P * -\alpha_i - \frac{C}{2}x_i$.

1.6 How Government Resources Affect Market Dominance

Up to this point the analysis are quite similar to Long and Wong with the exception of having two markets under the setting of sequential game. In this section, we will see how a city's resource constraint determines the extent of a market the city can capture. For simplicity, we only consider two cases here: (S,S) and (S,N). The first element in the parenthesis indicates the firm that capturing the entire financial market. For instance, if the first element is S, it means the financial market is dominated by financial firm S. Similar definition applies to the second element in the parenthesis, it indicates the distributor capturing the distribution market.

1.6.1 In the case of Financial Market

Analysis of Case (S,S)

In this case, both the financial and distribution markets are captured by the firms from city S. This is exactly identical to previous case in which the firms located in city N are so incompetent that the firms in city S are monopolistic service providers in their respective markets. Therefore, the equilibrium amounts of infrastructures built by government S are:

$$I_S^D = R_S\left(\frac{\psi_S^F}{\psi_S^F + \psi_S^D}\right) \tag{1.33}$$

$$I_S^F = R_S\left(\frac{\psi_S^D}{\psi_S^F + \psi_S^D}\right) \tag{1.34}$$

Accordingly, the firms' costs of providing services are:

$$C_S^D = \overline{C_S^D} - \frac{\psi_S^D}{2}\left(\frac{R_S\psi_S^F}{\psi_S^F + \psi_S^D}\right)^2 \tag{1.35}$$

$$C_S^F = \overline{C_S^F} - \frac{\psi_S^F}{2}\left(\frac{R_S\psi_S^D}{\psi_S^F + \psi_S^D}\right)^2 \tag{1.36}$$

Now recall that in the previous section that financial firm S will capture the entire financial market if and only if:

$$C_N^F \geqslant C_S^F + \frac{3}{2}C \tag{1.37}$$

Let us assume that city N follows a similar relationship between the amount of infrastructures and costs of providing services, that is:

$$C_N^j = \overline{C_N^j} - \frac{\psi_N^j\left(I_N^j\right)^2}{2}$$

j can be either F or D, depending on the market this cost associated with.

In order to guarantee the market dominance of financial firm S, it must be the case that even if city N devotes all its resources to build financial infrastructure, financial firm S

remains competitive enough so that condition (1.37) continues to hold. If city N dedicates all resources to build financial infrastructures, that is, $R_N = I_N^F$, then C_N^F becomes:

$$C_N^F = \overline{C_N^F} - \frac{\psi_N^F (R_N)^2}{2} \tag{1.38}$$

Substituting equations (1.36) and (1.38) into condition (1.37), we have

$$\overline{C_N^F} - \frac{\psi_N^F (R_N)^2}{2} \geqslant \overline{C_S^F} - \frac{\psi_S^F}{2} \left(\frac{R_S \psi_S^D}{\psi_S^F + \psi_S^D} \right)^2 + \frac{3}{2} C$$

Rearrange terms yields

$$R_S^2 \geqslant \left(\frac{2}{\psi_S^F} \right) \left(\frac{\psi_S^F + \psi_S^D}{\psi_S^D} \right)^2 \left(\overline{C_S^F} - \overline{C_N^F} + \frac{3}{2} C \right) + \frac{\psi_N^F}{\psi_S^F} \left(\frac{\psi_S^F + \psi_S^D}{\psi_S^D} \right)^2 R_N^2 \tag{1.39}$$

Condition (1.39) constitutes the sufficient condition under which the firms from city S capture both of the two markets. There are several explanations for this to happen. It may happen when (i) city S has more resources than city N, (ii) the financial infrastructure of city S is more effective than that of city N, or (iii) city S has an advantage on initial service cost, $C_S{}^F$, than the initial service cost of city N, C_N^F.

By symmetry, the financial and distribution firms in city N capture their respective markets entirely if:

$$R_N^2 \geqslant \left(\frac{2}{\psi_N^F} \right) \left(\frac{\psi_N^F + \psi_N^D}{\psi_N^D} \right)^2 \left(\overline{C_N^F} - \overline{C_S^F} + \frac{3}{2} C \right) + \frac{\psi_S^F}{\psi_N^F} \left(\frac{\psi_N^F + \psi_N^D}{\psi_N^D} \right)^2 R_S^2 \tag{1.40}$$

The above findings permit us to establish the following proposition:

Proposition 1. *If the relationship between the amount of resources(R_S^2) city S has and the amount of resources(R_N^2) city N has satisfies condition (1.39), then the financial and distribution firms located in city S capture entirely their respective markets, with the consequence that the firms located at city N capture none of the two markets. Contrarily, if the relationship between R_S^2 and R_N^2 satisfies condition (1.40), then the financial and distribution firms located at city N capture entirely their respective markets, with the consequence that firms located at city S capture none of the markets.*

Analysis of Case (S,N)

Now we consider an alternative case in which financial firm S captures the financial market and distributor N captures the distribution market. Given that the distribution market is dominated by distributor N, there is no incentive for government S to invest any resources on infrastructures that would help its distributor to lower operation cost. Instead, government S decides to allocate all the available resources to build financial infrastructures, i.e., $R_S = I_S^F$. We can justify this by reasoning through backward induction. Imagine that government S has to determine the amount of resources that should allocate to build infrastructure for distribution firm. By backward induction government S starts the game backward and examines the third stage first. Given the parameters government S knows it is hopeless in distribution market and therefore, its optimal strategy then is to spend all its available resources to build financial infrastructure instead. However, by looking at the third stage it is unclear whether government N will spend all its resources to build infrastructure for its distribution firm. Therefore, for financial firm S to dominate the financial market, it must be the case in which even if government N spends all its resources on enhancing financial infrastructures, financial firm S remains competitive enough to capture the entire financial market.

By substituting equation (1.38) into condition (1.37) and replacing I_S^F with R_S, we have

$$\overline{C_N^F} - \frac{\psi_N^F \left(R_N\right)^2}{2} \geqslant \overline{C_S^F} - \frac{\psi_S^F}{2} R_S^2 + \frac{3}{2} C$$

After rearranging terms, we obtain

$$R_S^2 \geqslant \left(\frac{2}{\psi_S^F}\right)\left(\overline{C_S^F} - \overline{C_N^F} + \frac{3}{2}C\right) + \frac{\psi_N^F}{\psi_S^F}R_N^2 \tag{1.41}$$

By symmetry, financial firm N captures the entire financial market whereas distributor S captures the entire distribution market if:

$$R_N^2 \geqslant \left(\frac{2}{\psi_N^F}\right)\left(\overline{C_N^F} - \overline{C_S^F} + \frac{3}{2}C\right) + \frac{\psi_S^F}{\psi_N^F}R_S^2 \tag{1.42}$$

We summarize the above conditions and combine them with Proposition 2:

Proposition 2: *If condition (1.41) holds, financial firm S captures the entire financial market. Furthermore, in the case in which firms located at city S capture both the financial and distribution markets, condition (1.39) must hold. Contrary, if condition (1.42) holds, financial firm N captures the entire financial market. Furthermore, condition (1.40) must hold for firms located at city N to capture both the financial and distribution markets. Finally, if both conditions (1.41) and (1.42) fail to hold, then no one is able to dominate the financial market.*

Figure 1.1 presents a graphical illustration of the findings shown above. Line AB represents condition (1.40). For any distribution of resources (R_S^2, R_N^2) that lies to the right of line AB, financial firm S captures the entire financial market. Line CD represents condition (1.39). For any distribution of resources (R_S^2, R_N^2) that lies to the right of line CD, both firms from city S capture their respective markets entirely.

The next case is when financial firm N dominates the financial market. Condition (1.40) and condition (1.42) are used to draw line GH and line EF, respectively. For any allocation of resources (R_S^2, R_N^2) that lies to the left of line EF, financial firm N captures the entire financial market. In addition, if such a resource distribution also lies to the left of line GH, both the financial and distribution markets are dominated by the firms from city N.

Finally, for any distribution of resources (R_S^2, R_N^2) that lies between line AB and line EF, no single firm is able to capture the entire financial market. In other words, this is the area where the financial market is shared by the two cities together.

Figure 1.1: *The distribution of government resources affects market dominance*

R_N^2

City N
dominates
both
markets

H

F

G

No single firm
dominates the
financial market

B

E

D

City S dominates
both markets

R_S^2

0

A

C
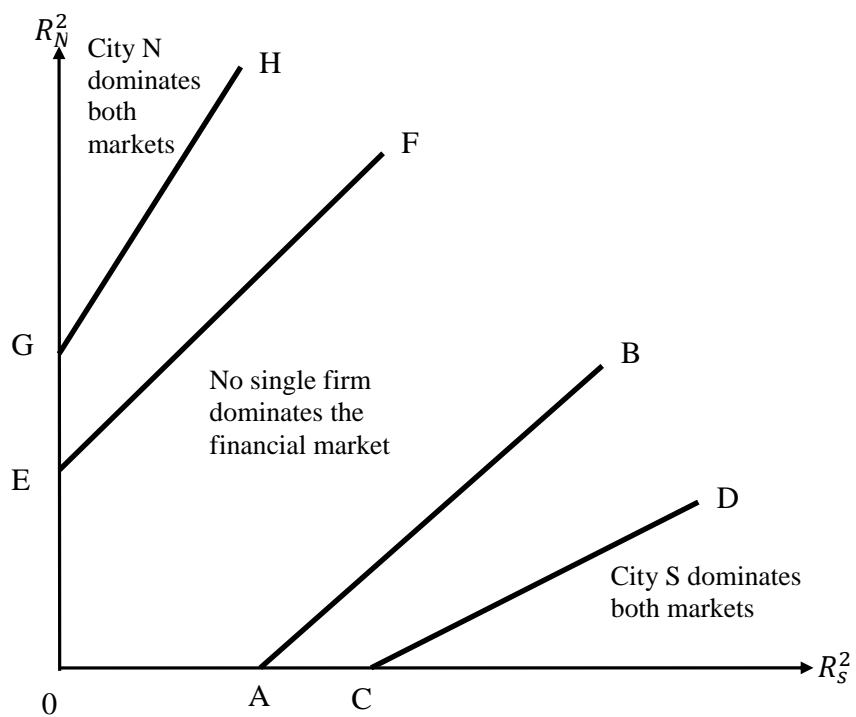
The Effectiveness of Financial Infrastructure

It is instructive to investigate how the effectiveness of financial infrastructure in a city affects its market dominance of the financial sector. By symmetry, we focus on the case for city S. If the city's financial infrastructure becomes more effective, other things being equal, the value of ψ_S^F increases. This case may emerge when, for instance, city S has a new policymaker who is more capable than his predecessor. Intuitively, an increase in effectiveness of financial infrastructure tends to enhance financial firm S in capturing the financial market. Looking at condition (1.41), larger ψ_S^F makes the RHS of the inequality smaller. In other words, it becomes relatively easier for city S to dominate the financial market. Figure 1.2 shows a graphical interpretation of this finding. An increase in ψ_S^F shifts the line AB to A'B', with the result that the area between AB and A'B' indicates the gains from this improvement.

However, it is unclear whether an increase in ψ_S^F will also help city S to capture the distribution market. On one hand, through equation (1.33), an increase in ψ_S^F increases I_S^D. As a consequence, distributor S becomes more competitive. On the other hand, the increase in ψ_S^F reduces the amount of financial infrastructure that government S would built. It makes condition (1.39) more difficult to satisfy. In words, it reduces the chance of city S to capture both the financial and distribution markets. The net result depends on the relationship between ψ_S^F and ψ_S^D:

1. If $\psi_S^D > \psi_S^F$, an increase in ψ_S^F improves the chances of city S to capture both markets;

2. If $\psi_S^D < \psi_S^F$, an increase in ψ_S^F reduces the chances of city S to capture both markets;

3. If $\psi_S^D = \psi_S^F$, an increase in ψ_S^F has no effect on capturing the distribution market.

   We summarize the findings in the following proposition:

Figure 1.2: *An improvement in financial infrastructure enhances market dominance of the financial sector*
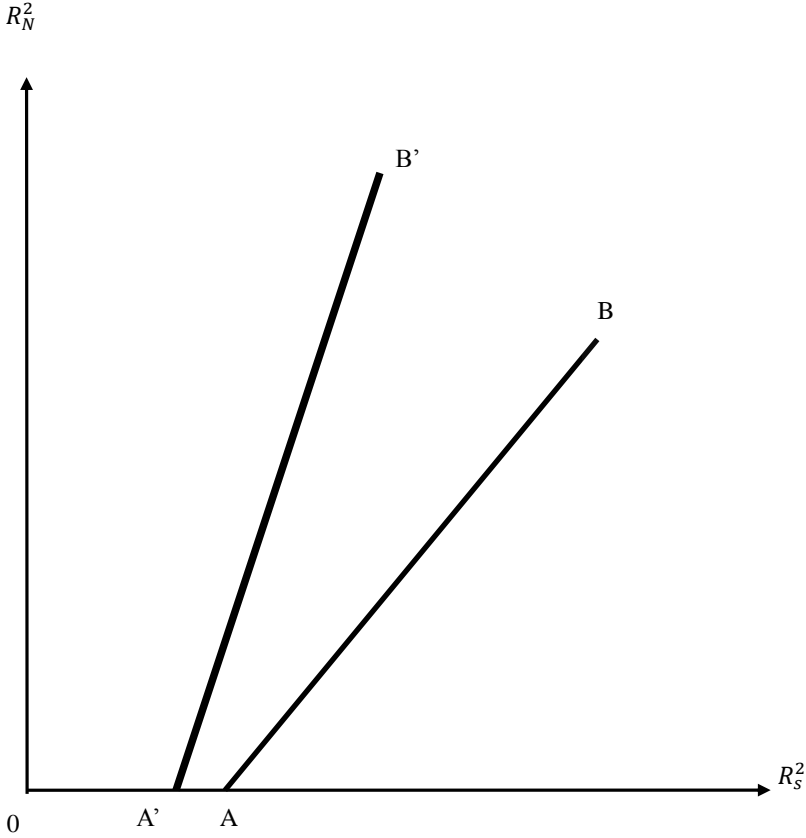
Proposition 3: *: An increase in ψ_i^F improves the chance of city i to capture the entire financial market. However, it is ambiguous whether this improvement will also help city i to capture the distribution market or not.*

Corollary 3: An increase in ψ_i^F, $i \neq j$, reduces the chance of city i to capture the entire financial market.

We can prove Corollary 3 easily by analyzing the improvement of ψ_S^F in city N's prospective.

Relative Effectiveness of financial infrastructure

Next, we wish to examine how relative effectiveness of financial infrastructure influences the dominance of financial market. Suppose $\frac{\psi_N^F}{\psi_S^F}$ goes down. This decrease may be caused by an increase in ψ_S^F, a decrease in ψ_N^F, or both. If it happens because of an increase in ψ_S^F or simultaneous changes in both ψ_S^F and ψ_N^F, we have from condition (1.41) that Figure 1.2 can be used to show how this change increases the competiveness of financial firm S. On the other hand, if this happens due to a decrease in ψ_N^F, then the graphical representation is slightly different. In this case the slope becomes steeper but the intercept remains unchanged. Figure 3 illustrates this change. Recall that condition (1.41) is used to graph line AB. With this increase the line switches to AB'. As in the previous case, the area between line AB and line AB' represents gains from the improvement on relative effectiveness of financial infrastructure.

However, it is ambiguous whether this improvement will also assist distribution firm S in capturing the distribution market or not. Proposition 3 applies if this improvement is caused by increase in ψ_S^F or simultaneous changes in both ψ_S^F and ψ_N^F. If this improvement is caused by reduction in ψ_N^F, we have from condition (1.39) that this reduction certainly helps firms in city S to capture both markets. We can summarize the above result:

Figure 1.3: *An improvement in the relative effectiveness of financial infrastructure enhances market dominance of the financial sector*
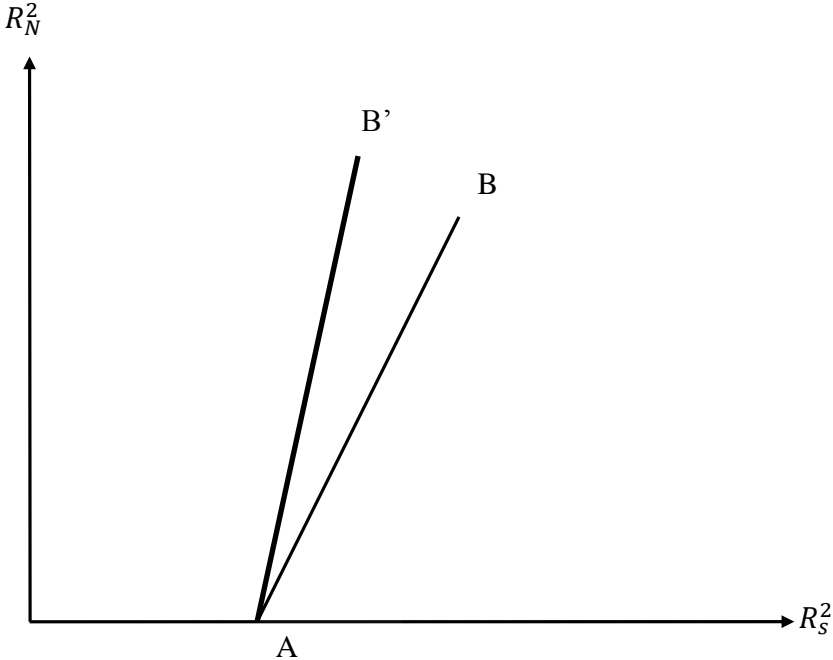
Proposition 4: *An improvement (A reduction) in the relative effectiveness of financial infrastructure certainly helps (prevents) city i to capture the financial market. Whether the magnitude of this increase (decrease) will enhance market dominance depends on the sources of this improvement. Nevertheless, it is ambiguous how changes in relative effectiveness affect the market dominance of the distribution sector.*

Changes in the Initial Cost Differential

In this section we examine how changes in initial cost affect the financial market. It is reasonable to assume that initial cost is positively related to the history of the city. Long established cities certainly have lower initial costs than those new established cities. Therefore, one can argue that by examining initial cost differential we can evaluate how history of the city affecting the financial market.

We continue to use city S as an example and suppose that the initial cost differential, $\overline{C_S^F} - \overline{C_N^F}$, decreases. It follows from equation (1.41) that this decrease causes a parallel shift in line AB to the left as shown in Figure 1.4. This change in initial cost differential helps financial firm S to serve more investors in the financial market. The gain in market share is measured by the area between line AB and line A'B'. Moreover, an investigation of equation (1.39) reveals that this decrease in the initial cost differential also helps distributor S to capture the distribution market. Since the term on the right hand side of equation (1.39) becomes smaller, it is more likely that the inequality condition is satisfied. We can summarize in the following proposition.

Proposition 5: *A decrease (An increase) in initial cost differential of providing services helps (prevents) city i to capture both the financial and distribution markets.*

1.6.2 In this case of Distribution Market

In this section we examine how government resource constraints affect the distribution market. Given that we consider a three-stage sequential game, we assume that financial firm S captures the entire financial market. That is, equation (1.41) hold.[3] As in Section 6.1, we consider only two cases: (S,S) and (S,N).

---

[3]Cases of partial financial market are ignored due to the complexity of mathematics.

Figure 1.4: *A decrease in initial cost differentical of providing services enhances market dominance of both sector*
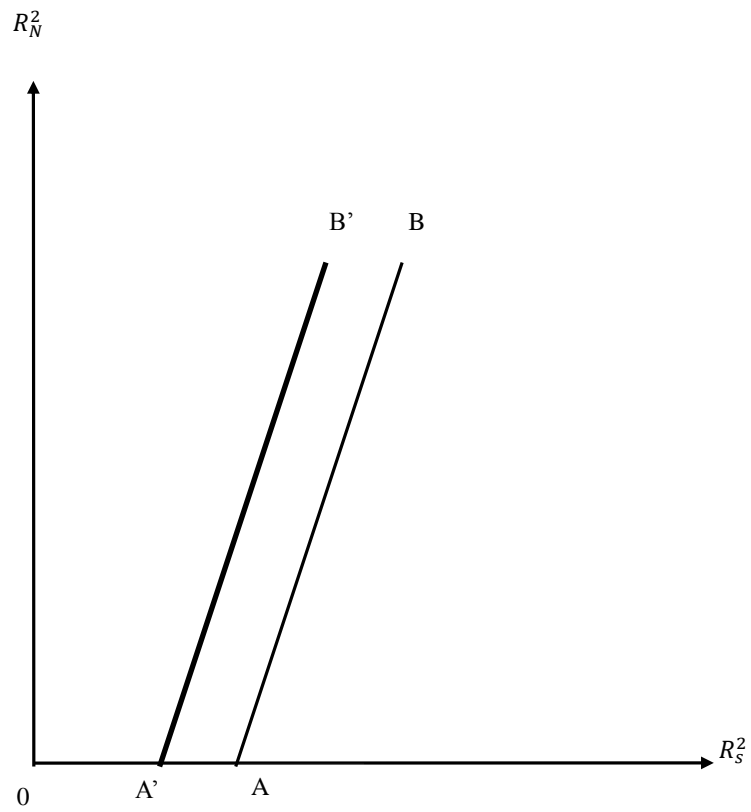
Analysis of case (S,S)

This is the case where firms in city S dominate both the financial and distribution markets. This case is identical to the one in which the firms in city S have monopolistic power over their respective markets. As a result, the amount of distribution infrastructure built by government S is identical to that in equation (1.28):

$$I_S^F = \frac{R_S \psi_S^D}{\psi_S^F + \psi_S^D} \tag{1.43}$$

The equilibrium service cost of distributor S is:

$$C_S^D = \overline{C_S^D} - \frac{\psi_S^D}{2} \left(\frac{R_S \psi_S^F}{\psi_S^F + \psi_S^D} \right)^2 \tag{1.44}$$

Now recall that in analyzing duopoly in the distribution market, firms in city S capture both the financial and distribution markets if and only if:

$$C_N^D \geq C_S^D + \frac{3}{2}b \tag{1.45}$$

For city S to dominate both markets, it must be the case that even if city N devotes all its available resources to build distribution infrastructure, distributor S remains competitive enough to capture the entire market. Substituting $R_N = I_N^D$ into C_N^D yields

$$C_N^D = \overline{C_N^D} - \frac{\psi_N^D (R_N)^2}{2} \tag{1.46}$$

Next, substituting equations (1.44) and (1.46) into equation (1.45), we get:

$$\overline{C_N^D} - \frac{\psi_N^D (R_N)^2}{2} \geq \overline{C_S^D} - \frac{\psi_S^D}{2} \left(\frac{R_S \psi_S^F}{\psi_S^F + \psi_S^D} \right)^2 + \frac{3}{2}b$$

Rearranging terms yields

$$R_S^2 \geqslant \left(\frac{2}{\psi_S^D} \right) \left(\frac{\psi_S^F + \psi_S^D}{\psi_S^F} \right)^2 \left(\overline{C_S^D} - \overline{C_N^D} + \frac{3}{2}b \right) + \frac{\psi_N^D}{\psi_S^D} \left(\frac{\psi_S^F + \psi_S^D}{\psi_S^F} \right)^2 R_N^2 \tag{1.47}$$

If firms in city S wish to dominate both markets, condition (1.47) has to hold. In addition recall from last section that condition (1.39) also has to hold for city S to capture both markets. Therefore, given the city of S capturing the financial market, both condition (1.39) and condition (1.47) have to hold for city of S to capture the distribution market as well.

Similarly, suppose city N captures the entire financial market, the following condition must hold:

$$R_N^2 \geqslant \left(\frac{2}{\psi_N^D}\right)\left(\frac{\psi_N^F + \psi_N^D}{\psi_N^F}\right)^2 \left(\overline{C_N^D} - \overline{C_S^D} + \frac{3}{2}b\right) + \frac{\psi_S^D}{\psi_N^D}\left(\frac{\psi_N^F + \psi_N^D}{\psi_N^F}\right)^2 R_S^2 \tag{1.48}$$

Once again, we know from the previous section condition (1.40) has to hold also for city N to capture both markets. We can summarize in the following proposition:

Proposition 6: *Suppose condition (1.41) holds and city S captures the entire financial market, city S will also capture the distribution market if conditions (1.39) and (1.47) are satisfied. Similarly, assuming that condition (1.42) holds and city N captures the entire financial market, city N will also capture the distribution market if conditions (1.40) and (1.48) are satisfied.*

Analysis of case (S,N)

This is the case in which city S dominates the financial market but city N dominates the distribution market. By backward induction we know that city S's optimal strategy is not to invest any infrastructure in the distribution market, i.e. $I_S^D = 0$ and $C_S^D = \overline{C_S^D}$. What about city N? By just observing the game in the third stage it is unclear what government N should do. Thus government N will move one stage backward to second stage. Once government N observes the second stage, it realizes that there is no hope to compete in the financial market. As a result, government N will allocate all its available resources to improving distribution infrastructure, that is, $I_N^D = R_N^2$. Given this condition, we have

$$C_N^D = \overline{C_N^D} - \frac{\psi_N^D(R_N)^2}{2} \tag{1.49}$$

It follows from the analysis in the section of "duopoly in distribution," case (S,N) emerges if and only if:

$$C_S^D \geq C_N^D + \frac{3}{2}b \tag{1.50}$$

Replacing C_S^D with $\overline{C_S^D}$ and substituting equation (1.49) into equation (1.50), we have

$$\overline{C_S^D} \geq \overline{C_S^D} - \frac{\psi_N^D (R_N)^2}{2} + \frac{3}{2}b$$

Rearranging terms yields

$$R_N^2 \geq \frac{2}{\psi_N^D}\left(\overline{C_N^D} - \overline{C_S^D}\right) + \frac{3}{\psi_N^D}b \tag{1.51}$$

Thus, the above equation constitutes the sufficient condition under which case (S,N) arises. In addition, given that financial firm S dominates the financial market, condition (1.413) also has to satisfy. There are two possibilities:

1. If $\frac{2}{\psi_N^D}\left(\overline{C_N^D} - \overline{C_S^D}\right) + \frac{3}{\psi_N^D}b > 0$, the terms on the right hand side of equation (1.51) becomes non-negative. Figure 1.5 illustrates this case. Line AB represents condition (1.41), line CD represents a line that satisfies both conditions (1.40) and (1.48). Line EF represents condition (1.51). We know that case (S,N) must satisfy both condition (1.41) and condition (1.51). The area that satisfies both conditions is the area that bounds between BE and DF, and bounds below by EF. In addition, the area bound above by EF and bound between BE and DF represents the case in which financial firm S dominates the financial market, but firm captures the entire distribution market.

2. If $\frac{2}{\psi_N^D}\left(\overline{C_N^D} - \overline{C_S^D}\right) + \frac{3}{\psi_N^D}b \leq 0$. In this case the RHS of condition (1.51) is non-positive. Figure 1.6 illustrates this case. The obvious difference between this case and the previous case is the non-existence of a partial distribution market. In this case, the distribution market is dominated by either distributor S or distributor N.

The above findings allow us to establish

Figure 1.5: *A case in which the partial distribution market exists.*
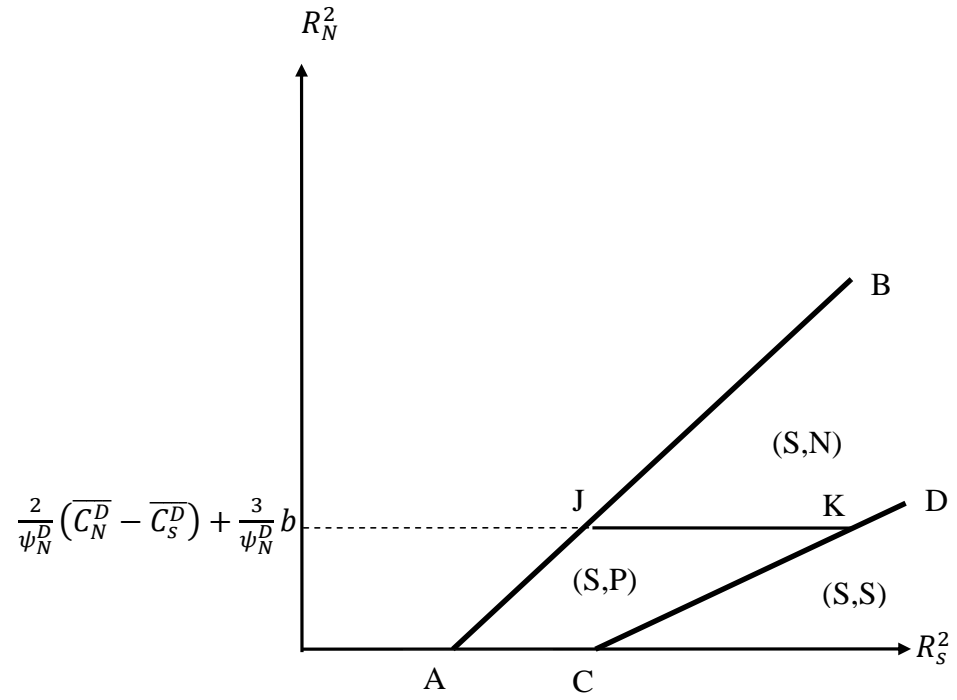
Figure 1.6: *A case in which the distribution market capture entirely by city N.*
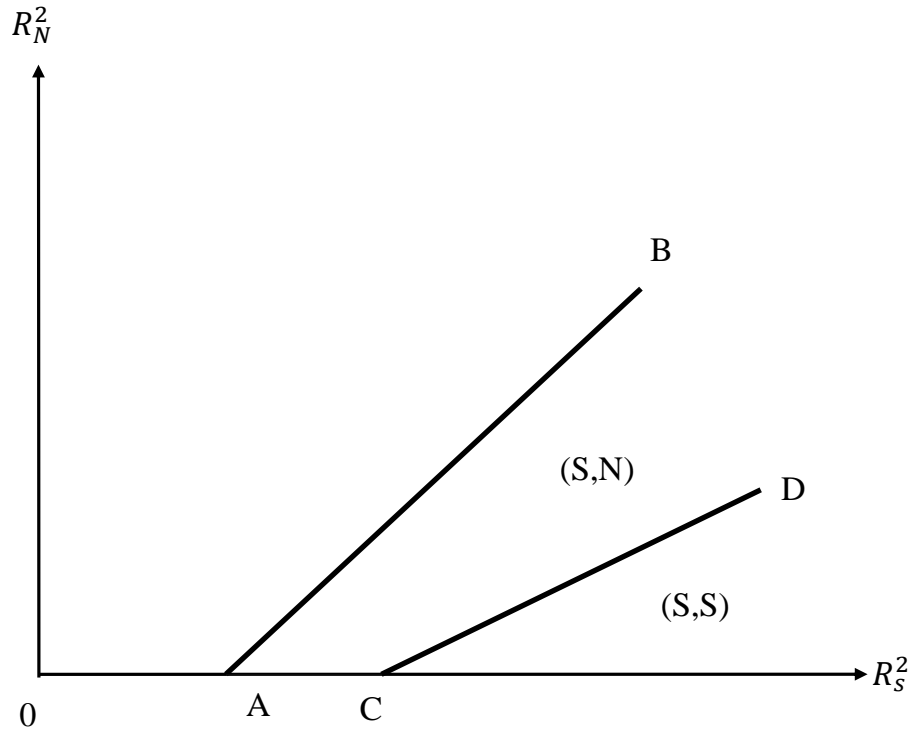
Proposition 7S: *Suppose city S dominates the financial market and condition (1.41) holds:*

If both the sufficient conditions in (1.39) and (1.47) hold, city S captures both the financial and distribution markets.

If either condition (1.41) or condition (1.47), or both, fail to hold, then there are two cases:

1. *If $\frac{2}{\psi_N^D}\left(\overline{C_N^D} - \overline{C_S^D}\right) \leq \frac{3}{\psi_N^D}b$, partial distribution market does not exist. In other words, the distribution market is dominated by distributor N;*

2. *If $\frac{2}{\psi_N^D}\left(\overline{C_N^D} - \overline{C_S^D}\right) > \frac{3}{\psi_N^D}b$, then partial distribution market exists.*

By symmetry, we summarize the situations in which city N dominates the entire financial market.

Proposition 7N: *Suppose city S dominates the financial market and condition (1.42) holds:*

If both the sufficient conditions in (1.40) and (1.48) hold, city S captures both the financial and distribution markets.

If both condition (1.40) or condition (1.48), or both fail to hold, then there are two cases:

1. *If $\frac{2}{\psi_S^D}\left(\overline{C_S^D} - \overline{C_N^D}\right) \leq \frac{3}{\psi_S^D}b$, partial distribution market does not exist. In other words, the distribution market is dominated by distributor S;*

2. *If $\frac{2}{\psi_S^D}\left(\overline{C_S^D} - \overline{C_N^D}\right) > \frac{3}{\psi_S^D}b$, then partial distribution market exists.*

1.7   Concluding Remarks

In this paper we extend the model of Long and Wong (2009) to allow for a multiple dimension analysis where two cities compete to become the distribution headquarter and the financial center. We show that if the resources of one city is relatively abundant than the other city, this city tends to dominate both the distribution and financial markets. This conclusion is reached under the assumption that city governments efficiently use their limited resources to

undertake cost-reducing investments in infrastructures. Given that the amount of resources a city has may reflect the size and the prosperity of the city, our analysis implies that larger and wealthier cities are more likely to become both the distribution and financial centers. We observe this phenomenon in the real world, for instance, Singapore is the main financial market in South East Asia and Manila is not.

We further show that an improvement in infrastructures enhances the competitiveness of either the financial or the distribution sector. This result is not limited to the cases that involve physical infrastructures such as buildings. It also applies to various situations that involve laws and/or economic systems cities have. Better legal system helps improve the competitiveness of local firms in international markets by infusing assurances and trusts to them, as we frequently observed in real world. One of the common features that major financial cities such as New York, Hong Kong, Tokyo and London shares is a sophisticated legal system. Our simple model provides a possible explanation as to why each of these cities is capable of becoming a major player in the financial industry, either regionally or globally.

Some caveats and hence potentially interesting extensions of the present model should be mentioned. First, we assume in our simple analysis that two competing cities share similar geographical background. Certainly this is far away from reality. Geographical location affects the types of cities and the competitiveness of firms located within the cities. It is not a coincidence that almost all majority cities in the world today locate along the coastline. One can argue that the initial costs in the model reflect some degrees of geographic advantage but they are not sufficient enough to explain thoroughly how location affects the market outcome. A more complete framework of a spatial competition should be considered in the future research. Another possible extension is to incorporate production technology into the analysis. In this case, the quality of labor and the return on capital may constitute important factors in determining whether cities can be major players in their respective markets. Economies of scale in production many also play a crucial role in affecting the

equilibrium outcome of market competition.

Chapter 2

Income Distribution and the Composition of Imports

2.1 Introduction.

Traditional trade theories usually focus on the technology of producing tradable goods. For example, Ricardian models assume the differences in the method of production and emphasize the importance of the comparative advantages. Heckscher-Ohlin models argue that countries should produce goods that best fit to their factor endowments. These supply based theories dominated the trade literature until the introduction of Linder's hypothesis. Linder (1961) emphasizes the importance of the internal demand within the home country. Goods are tailored specifically to satisfy the consumer preferences of domestic consumers, which are likely determined household income levels. Further, these are goods most likely exported to countries with similar consumer preferences. He concludes that countries that share similar income levels will trade more often. We extend this analysis to examine the pattern of import composition across varying income distributions. While the use of per capita income measures are prevalent in the existing literature, only a limited number of studies explore the relationship between income distributions and trade composition. This specific relationship is the main objective in this paper.

We argue that importing countries with similar income distributions will share similar import patterns. To explore this idea, we relax the typical assumption of homothetic pref-

erences. While most modern trade models assume that homothetic preferences, such preferences do not allow us to directly examine the role of consumption behavior by households of varying income levels. To address this concern, we develop an import similarity index, which allows us to compare the composition of imports from a source country across countries. Likewise, we seek to address the role of income distributions in determining the composition of imports and, therefore, apply the concept of "market overlap", an idea advocated by Bohman and Nilsson (2007b), to characterize similarities in the income distributions of specific countries. We provide empirical evidence to support the notion that countries with similar income distributions display similar import patterns. We also conduct robustness checks that verify that are results hold across the middle- and high-income subsamples of countries as well as within the differentiated and reference priced goods categories.

2.2 Related Literature

Prior to the recent decade, several attempts have been made to verify the validity of Linder's hypothesis by using the income per capita at national level. They use income per capita to model the representative preferences and, thus, the demand for countries. Kohlhagen (1977) uses real aggregate private consumption as proxy for demand in his paper. He concludes that differences in demand between countries partially explained the amount of trade patterns in his countries sample. More recently, several models built upon the relationship between incomes, preferences and characteristics of the exported goods. People in high income countries will spend large portion of their incomes on high quality goods. As a result, rich countries tend to produce high quality goods and export them to other rich countries. Hallak (2006) uses export prices to represent the quality of exported goods and associate it with importer income per capita to show that quality is indeed a good indicator of the direction of trade flows. In another paper, Hallak (2010) argues that the sectoral data should be used to test the Linder hypothesis. He shows that the reason why so many empirical papers fail to find evidence to support the Linder hypothesis is due to their use of aggregate data.

This aggregation bias occurs because different sectors are correlated with income per capita differently. Using importer's income per capita and exporter's income per capita as indicator of quality of goods, he shows that sectoral data will provide robust empirical evidence to support the Linder hypothesis despite the fact that tests utilizing aggregate data will fail.

Until recently, homothetic preferences have been assumed when applying gravity equations to access the relationship between the trade flows and income. Matsuyama (2000) provides theoretical model under the assumption of nonhomothetic preferences. The South (North) has a comparative advantage in producing goods with low (high) income elasticites of demand. He shows that redistributive policies by the South can improve their terms of trade at the expense of the North. Dalgin et al. (2008) make a convincing case to show why nonhomothetic preference and income distribution should also be included when estimating trade flows. First, nonhomothetic preferences are more realistic to the real world and are consistent with studies of the income elasticity of demand (see Hunter (1991), Linder Hunter and Markusen (1988) and Bohman and Nilsson (2007a)). Second, given nonhomothetic preferences, the use of income distributions becomes necessary to capture the demand for goods and the resulting trade patterns.

Given the premise of nonhomothetic preferences, Dalgin et al. (2008) find that as inequality increases in importing countries, they tend to import more from the rich countries, as those rich countries most likely produce and export luxury goods. On the other hand, imports from poor countries tends to fall as inequality increases, due to the fact that poor countries tend to produce necessities. Their results are closely related to the Linder's theory, as income distributions can be use to characterize the consumer preferences in the import country. As the income distributions in an importing country changes, demand of goods also changes according to the characteristics of that specific goods. Another way to express this is that preferences tend to converge as the level of incomes and income distributions between import country and export country converge. As a result, these two countries trade more often, the main conclusion that Linder hypothesis draw.

While Dalgin et al. (2008) use Gini coefficients as an index of income dispersion, Bohman and Nilsson (2007b) use the concept of "market overlap" to characterize the income distribution. In their paper they study the relationships between overlapping markets of trading partners and their respective trade flows. Their underlying assumption is that countries with similar income levels share similar preferences or tastes toward particular goods. Comparing the income distributions of both the exporting and importing countries allow them to construct the "market overlap" of the two trading countries. This "market overlap" represents the overlapping demand that these two countries share. This provide another method to verify the Linder hypothesis. By using this concept they are able to find empirical evidence to support the notion that countries with larger overlapping markets trade more often.

Fajgelbaum et al. (2009) also develop a theoretical model with the assumption of a nonhomothetic utility function. Furthermore, they incorporate the quality and the variety of the product into consumer preferences. With nonhomothetic demand, they are able to depict the linkage between the income distributions of the trading countries and the pattern of its trade in differentiated products. They show high income countries export high quality goods and import low quality goods. Also, poorer households in the richer countries and richer households in the poorer countries stand to gain most if trade liberalization occurs.

While the Linder hypothesis postulates the trade flow of trading countries using income distributions, it does not provide any measure about trade pattern. Finger and Kreinin (1979) propose an index that will capture the similarity of exporting patterns of trading countries. Since then, several modifications of this index have been used to study the pattern of trade (See Lewis and Webster (2001) and Zheng and Qi (2007)). Schott (2008) uses this index to study the overlapping export patterns between China and OECD countries. He find that while the export patterns between China and OECD countries become similar, the export prices of China are much lower comparing to the OECD counterparts. We discuss our modified version of this index that can be used to contrast the composition of imports in Section 2.4.

2.3 Theory

To address the difference in import composition across countries of varying income distributions, we assume preferences are nonhomothetic in our theoretical model. This allows the composition of consumption bundles to differ across individuals of different income levels, such as the relative consumption of luxury goods versus necessities. We follow the model of Dalgin et al. (2008) in expressing the demand for good g by household h as

$$D_{gh} = D_{gh}(p_g, I_h), \tag{2.1}$$

where I_h is household income and p_g is the price of the good. As Dalgin et al. (2008) note, aggregate demand for good g in a model with a representative consumer, D_g, typically utilizes per capital income:

$$D_g = D_g(p_g, \bar{I}),$$

where per capital income for N households is $\bar{I} = \frac{1}{N}\sum_{i=1}^{N} I_i$. However, they suggest that such an assumption is inappropriate in the case of nonhomothetic preferences, and thus aggregate demand should be considered a function of the income of each household in the country:

$$D_g = D_g(p_g, I_1, I_2, ..., I_N). \tag{2.2}$$

Our objective is to directly address overlapping income distributions across importing countries.[1] Accordingly, we assume that (2.2) can be expressed with additive functions for country b as:

$$D_{gb} = \tilde{\beta}_{0b} + \tilde{\beta}_p p_{gb} + \Gamma_2(I_{1b}, I_{2b}, ..., I_{Nb}). \tag{2.3}$$

We now express the demand for good g as a share of expenditure and define the function embodying the distribution of income as

$$d_{gb} = \beta_{0b} + \beta_p p_{gb} + \gamma_2(i_{1b}, i_{2b}, ..., i_{Qb}, \bar{I}), \tag{2.4}$$

---

[1]Dalgin et al. (2008) make use of a Gini coefficient to proxy income equality in their analysis of the trade of luxuries and necessities.

where $i_{1b}...i_{Qb}$ denote the respective income shares of households in each income quantile, which we discuss in the next section, and β_{0b} and β_{pb} represent the respective scaled coefficients $\tilde{\beta}_{0b}$ and $\tilde{\beta}_{pb}$.

For any two countries, b and c, the difference in demand for good g can be expressed as:

$$d_{gb} - d_{gc} = (\beta_{0b} - \beta_{0c}) + (\beta_p p_{gc} - \beta_p p_{gc}) + (\gamma_2(i_{1c}, i_{2c}, ..., i_{Qc}, \bar{I}) - \gamma(i_{1c}, i_{2c}, ..., i_{Qc}, \bar{I})), \quad (2.5)$$

which highlights the notion that the differences in income distributions across countries, $(\gamma_2(i_{1c}, i_{2c}, ..., i_{Qc}, \bar{I}) - \gamma(i_{1c}, i_{2c}, ..., i_{Qc}, \bar{I}))$, lead to differences in the demand for good g. We make use of this concept as we link our theory of income distribution-based demand to the import behavior of countries of contrasting income distributions.

2.4   Empirical Methodology

Equation (2.5) states that differences in demand for good g for any two countries b and c are driven by their differences in income distributions. Conversely speaking, if their income distributions are similar, then their demand for good g should be similar. More generally, countries with similar income distributions should share similar import patterns. In this section, we introduce the use of import similarity index and the concept of "market overlap" to characterize import patterns of countries and income distributions of countries, respectively.

2.4.1   Import Similarity Index

Finger and Kreinin (1979) introduce an index that measures the similarity of *export* patterns between two countries. In this paper we adopt this idea and modify it to measure the similarity of *import* patterns of two countries. We call this the import similarity index (ISI). This index takes the form of :

$$ISI_{bc,a} = \sum_{s=1}^{S} min(Im_{ab,s}, Im_{ac,s}) \tag{2.6}$$

$ISI_{bc,a}$ denotes as the import similarity of two importing countries, b and c. Both countries import from third country a. S is the total number of sectors. $Im_{ac,s}$ denotes the share of trade in sector s to total trade from country a to country c. Therefore, this index measures the import similarity of two countries by their aggregate import shares importing from the same third country. ISI is bounded between 0 and 1, and we say two importing countries share similar import patterns if the ISI is close to 1.

However, there is a mathematical misspecification in this index. Imagine there are two countries, x and y and both of them do not import any goods from country a. Intuitively, they have the exact identical importing pattern, thus the index should be 1. However, according to formula (2.6) the ISI index would be zero, indicating the importing patterns of these two countries are totally different. We rectify this problem by assigning the ISI index to 1 every time this situation occurs. In addition, we conduct a robustness check in section 2.6.2 and the results suggests that even without changing the ISI to 1 for countries without trade, this does not substantially change the results.

2.4.2   Market Overlap and Income Distributions

We utilize the concept of market overlap, henceforth MO, which is a concept introduced by Bohman and Nilsson (2007b). Suppose there are two countries, i and j with income distribution $\theta_i(y)$ and $\theta_j(y)$, respectively. We assume income distributions are function of disposable income, y. We define the market overlap (MO_{ij}) of these two countries as:

$$MO_{ij} = \int_0^\infty min[\theta_i(y)\theta_j(y)]dy. \tag{2.7}$$

In other words, we integrate the density of the income distributions these two countries shared across all income levels. However, in order to apply this approach, quality data on countries' income distributions are necessary for estimations. Optimally, we would like to

have income data at household level so that we would know the exact income distributions. In reality, countries usually only provide income distributions either in term of quantiles or deciles. Thus, in order to obtain more accurate empirical results, we applied the kernel density estimation suggested by Sala-I-Martin (2006) to derive our income distributions for countries in our sample. Several papers have used kernel smoothing techniques to estimate the income distribution. Jenkins (1995) applies this technique to estimate the income distribution of the UK when investigating the hypothesis of shrinking middle class in the 1980s. Johnson (2000) uses the adaptive kernel method to estimate the income densities of each state in years 1948, 1963, 1978 and 1993 by using the respective US state-level income per capita. Recently, Dai and Sperlich (2010) use boundary bias correction kernel smoothing in estimating the world income distribution.

Let $x_1, x_2, \ldots x_n$ (n is 5 if quintiles and 10 if deciles) represent the income share of a particular country we want to estimate. Then the shape of the income distribution g with the kernel density estimation is:

$$g_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x - x_i}{h})j, \tag{2.8}$$

where

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}. \tag{2.9}$$

Equation (2.9) is the Gaussian function that we apply in kernel density estimation. The bandwidth of this estimation, h, serves as a smoothing parameter. Similar to Bohman and Nilsson (2007b), we smooth and partition the entire income distribution into two permilles with bins incremented by 100 U.S. dollars. We then normalize the kernel density so the total density equals to 1. Once we estimate the income distributions of all countries in our sample, we can apply equation (2.7) to derive the MO for all trading partners.

We will replace the demand differences in equation (2.5) with the ISI and the differences in income distributions with MO. Furthermore, we will corporate several factors that are commonly used in trade literature. In generally, we are interested in estimating:

$$ISI_{bc,a} = \psi_b + \psi_c + \varphi_a + \alpha distdiff_{bc,a} + \beta \mathbf{I_{bc,a}} + \gamma MO_{bc} + \eta_{bc,a} \qquad (2.10)$$

where $distdiff_{bc,a} = |ln(Distance_{ab}) - ln(Distance_{ac})|$— and $Distance_{ab}$ denotes the distance between country a and country b. $\mathbf{I_{bc,a}} = |\mathbf{I_{ab}} - \mathbf{I_{ac}}|$ and matrix $\mathbf{I}_{ab}$ contains other control variables of both countries, such as common languages and currencies. In addition, we include the fixed effect of the exporting country, φ_a, and the fixed effects of the importing countries, ψ_b and ψ_c, in our estimation. Equation (2.10) is the baseline estimation model for this paper. If our theory is correct, such that countries with similar income distribution share similar import patterns, then γ should be positive. Also, we expect α to be negative, because trade costs will play a larger role in trade composition as distances vary to a greater level. Thus, a negative α suggests that as the difference of the distance of each importers between the same exporting country increases the import patterns should be more dissimilar to each other. Our results for the above estimations are reported in section 2.6.

2.5 Data and Descriptive Statistics

2.5.1 Data

The bilateral trade flow data used in this study are from Feenstra et al. (2005). While these data cover the years 1962-2000, we limit our study to the year 2000. We aggregate these data to the three-digit SITC Revision 3 classification before we create the ISI and otherwise conduct the analysis. Lastly, we limit our sample to 60 countries, which are shown in Table 2.1.

The data for income distributions is obtained from the UNU-Wider World Income Inequality Database (2008), henceforth WIID. These data include information regarding the income share of individuals (or households) in a given quantile for a given country. When information regarding income share deciles in the year 2000 were available, we utilize these data in the kernel estimate used for the creation of MO. When decile information was not present, we utilized the quintile income information. If data were not present for the year

45

2000, we use data from nearby years. While the UNU-Wider WIID provides the information on the distribution of income, the data for per capital gross national income comes from the World Bank's World Development Indicators (2011).

Lastly, the data on the gravity variables used as controls in our estimation are from two sources. Data on distance, population, common borders, common languages, colonial relationships, and common currencies come from the Trade Protection and Production Database (Nicita and Olarreaga 2007). Data indicating whether two countries have a regional trade agreement were obtained from the CEPII Gravity Dataset (Head et al. 2010).

2.5.2   Descriptive Statistics

Market Overlap

In this section, we will provide basic statistics on the market overlap and import similarity index. By definition, MO represents where the income distributions of two countries overlap. Since there are 60 countries in our sample, there are 1700 (60×59) MOs. Figure 2.1 shows the size of MO across country pairs. In our sample, more than half of our country pairs have an MO of 50 percent or more. In addition, more than a quarter of the country pairs overlap their markets for about 80 percent. On the other hand, there are less than 25 percent of the country pairs where MO is less than 20 percent. Furthermore, MO should be negatively related to the log income differences, since MO represents the overlapping sections of the income distributions of two countries. Figure 2.2 plots the log income differences against MO. It is clear that log difference in income is negatively related to MO. That result is not surprising since if the income difference between two countries is small, their markets are more likely to overlap and vice versa. However, we also note that variations in MO do exist for country pairs with similar differences in per capita income. These deviations, and their impact on trade composition, are the key focus of this paper.

Import Similarity Index

The import similarity index (ISI) represents the common spending share of imported goods of two import countries importing from the third exporting country. There are 60 countries in our sample, resulting into 102660 ($60 \times 59 \times 58/2$) observations of ISI. Figure 2.3 represents the ISI and its percentage share. In our full sample, around 38 percent of the observations have the ISI that are less than 0.1 and around a quarter of our sample have the ISI that are higher than 0.4. These results are consistent with the notion that countries with different income level do not share similar import patterns. Therefore, we divide countries in our sample into three different country groups based on their GNI per capita: high income countries (GNI per capita¿12000), mid level income countries (12000¿GNI per capita¿2000) and low income countries (GNI per capita¡2000).

Figure 2.4 shows the ISI between high income countries. There are less than 13 percent of our observations where the ISI is less than 0.1 and around 50 percent of our observations have the ISI that are higher than 0.4. In contrast, figure 2.6 the ISI are severely skewed to the right, meaning the majority of these low income countries do not share similar import patterns between them. Figure 2.5 presents the case of middle income countries, the ISI are also skewed to the right, but not as severe as in the case of low income countries.

2.6 Results

2.6.1 Baseline results

Our empirical results are presented in this section. Equation (2.10) is our fully specified model:

$$ISI_{bc,a} = \psi_b + \psi_c + \varphi_a + \alpha distdiff_{bc,a} + \beta \mathbf{I_{bc,a}} + \gamma MO_{bc} + \eta_{bc,a} \qquad (2.11)$$

where $I_{bc,a}$ include several variables that are commonly used in trade literature: *gnidiff*, *rtadiff*, *contigdiff*, *comlangdiff*, *col45diff*, *comcurdiff* and *popdiff*. Their definitions are as

follows: $gnidiff = |log(gni_b) - log(gni_c)|$ where gni_i denotes as the gross national income per capita of country i, thus $gnidiff$ is the absolute difference of gross national income per capita between importing country b and importing country c. The dummy variable $rtadiff$ captures the trade agreement status of two importing countries, b and c relative to the same exporting country a. It is equal to one if only one of the two importing countries have a trade agreement with exporting country. It equals zero if both importing countries have the free trade agreement with the same exporting country or if both importing countries do not have the free trade agreement with the same exporting country. Other control variables are constructed similarly. The variable $contigdiff$ captures the geographical aspect of two importing countries. It equals one if only one of the two importing countries locates next to the same exporting country, and zero otherwise. The variable $comlangdiff$ equals one if only one of the importing countries and the exporting country use the same language, and zero otherwise. The dummy variable $col45diff$ represents the colonial relationship. It equals one if only one of the two importing countries has the colonial relationship with the exporting country prior to 1945. A dummy variable for currency, $comcurdiff$, equals one if only one of the importing countries and the exporting country use the identical currency in their respective countries. Finally, $popdiff$ controls for the difference in logged population between the two importing countries.

Table 2.2 reports the results of various versions of equation (2.10). In all four cases, we obtained the results that we expected. The estimated coefficients on MO are all positive and are statistically significant at the 1 percent level in all four cases, suggesting that as the mark overlap of two importing countries become larger, the import patterns of these two countries will become more similar. Furthermore, the results indicate that if the differences of income per capita become smaller, these two countries will share similar import pattern. Finally, our results are similar to other studies utilizing the gravity model: if two countries are further away from each other, their import patterns will be more dissimilar.

2.6.2 Robustness Checks

For our first robustness check, we want to examine the effect of countries with zero trade flows on our results. Therefore, our next empirical estimations we eliminate country-pairs with zero trade flow. We report our results in Table 2.3. We obtained similar results by using this subsample of the original data. The estimated coefficients on *MO*, *gnidiff* and *disdiff* still have the expected signs and are statistically significant at the 1% level. In addition, the R-squared increases more than 15 percent in all four cases. With these results we conclude that our method of handling countries with zero trade flow does not substantially influence our results.

2.6.3 Income Subsamples

As a second robustness test, we divide these 60 countries into three different income groups (as in section 5.2.2) and estimate equation (2.10) again. Results are reported in Table 2.4. In the case of high income countries, the results are similar to the previous subsection. As the income distributions of the importing countries become similar, the import patterns become similar as well. However, it is unclear how the differences in income per capita will affect the import patterns. When *MO* is absence in the estimation, we obtain the expected sign on differences in income per capita income. However, when *MO* is included in the estimation, *gnidiff* becomes positive and is statistically insignificant.

We obtain similar results for the mid-level income countries, *MO* is still positively related to ISI and are statistically significant in all four cases. Similarly, the estimated coefficients on differences in income per capita is negative with the absence of *MO* and positive with *MO* in the estimations. In addition, R squared decrease dramatically in all four cases.

From Table 2.4, we observe that the estimated coefficients on *MO* are all negative and statistically significant for low income countries. These results contradict our hypothesis that countries with similar income distributions should share similar import patterns. However, this hypothesis is based on the demand-side characteristics of importing countries, which

are largely influenced by differentiated goods and product quality. If low income countries import proportionally more homogeneous goods, then a greater amount of trade may be driven by such influences as domestic production abilities, transportation costs, and trade barriers.

We report the proportion of different type of imported goods for each of the three different country groups in Table 2.6. Notice that both high and middle income countries, import share of differentiated goods are above 50 percent. In comparison, in the case of the low income country, the import share of differentiated goods accounts for 38 percent only. In addition, the import of homogeneous goods is approximately 32 percent. The fact that the low income countries import less differentiated goods and more homogeneous goods among themselves. These results provide a plausible explanation as to the previously discussed results for lowon MO are negative, since the low income countries import much fewer differentiated goods than the high income countries and the mid level income countries. We further examine the role of goods classifications in our next robustness check.

2.6.4   Goods Classification Subsamples

Finally, we are interested to see how the types of goods affecting our results. Thus, we divide sectors into three different types of goods according to Rauch (1999) classification: homogeneous goods, differentiated goods and references goods. Homogeneous goods are goods that trade in the organized exchange, such as petroleum and gold. References goods are goods that have the reference prices, these are the prices that customers find on trade publications, such as chemicals. Differentiated goods are goods that belongs to neither cases.

Table 2.5 reports the results. While the results of differentiated goods and references goods are similar, the results of homogenous goods are quite different. In the cases of differentiated goods and references goods, the estimated coefficients on MO are positive and statistically significant. On the other hand, in the case of homogenous goods, even though the estimated coefficients on MO are positive, they are not statistically significant.

These results are consistent with Hallak (2006) and Hallak (2010). They state that quality is the center piece of the Linder hypothesis. Since homogenous goods are identical in terms of quality and variety, those goods are not produced to fit a specific set of preference and therefore, the Linder hypothesis is less applicable to these goods.

2.7 Conclusion

In this paper we study the relationship between the income distributions of importing countries and the composition of imports. We obtain income distributions and smooth the data using kernel estimates, which allow us to construct the market overlap of countries. These market overlaps represent the overlapping demand in given country pairs. This concept allow us to provide empirical evidence to show that income distributions and import patterns are closely related. We conclude that countries with similar income distributions share similar import patterns. When we separate our countries by their income level, income distribution still exerts a positive impact on the similarity of import patterns for all but low income countries. Finally, we incorporate the characteristics of goods into our analysis. We find that while the positive relationships between the income distributions and the import patterns still hold for differentiated goods and references goods, the same cannot be said about homogeneous goods. This final result is not suprising given that homogeneous goods are identical in terms of quality and variety.

Table 2.1: *List of Countries*

Argentina	Australia	Austria	Belarus	Belgium	Brazil
Bulgaria	Canada	Chile	China	Colombia	Costa Rica
Croatia	Czech Republic	Denmark	Ecuador	Egypt	Estonia
Finland	France	Germany	Greece	Hungary	Indonesia
Iran	Ireland	Israel	Italy	Japan	Kazakhstan
Latvia	Lithuania	Malaysia	Mexico	Netherlands	New Zealand
Norway	Pakistan	Peru	Philippines	Poland	Portugal
Russia	Slovakia	Slovenia	South Africa	South Korea	Spain
Sri Lanka	Sweden	Switzerland	Thailand	Tunisia	Turkey
Ukraine	United Kingdom	United States	Uruguay	Venezuela	Viet Nam

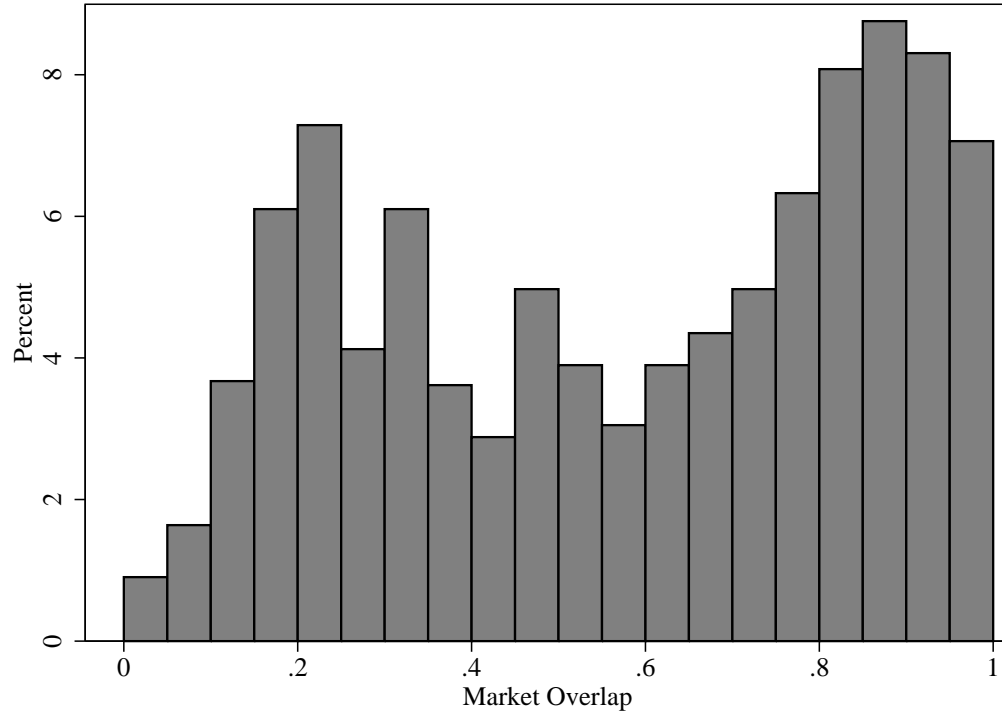**Figure 2.1**: *Distribution of Market Overlaps*

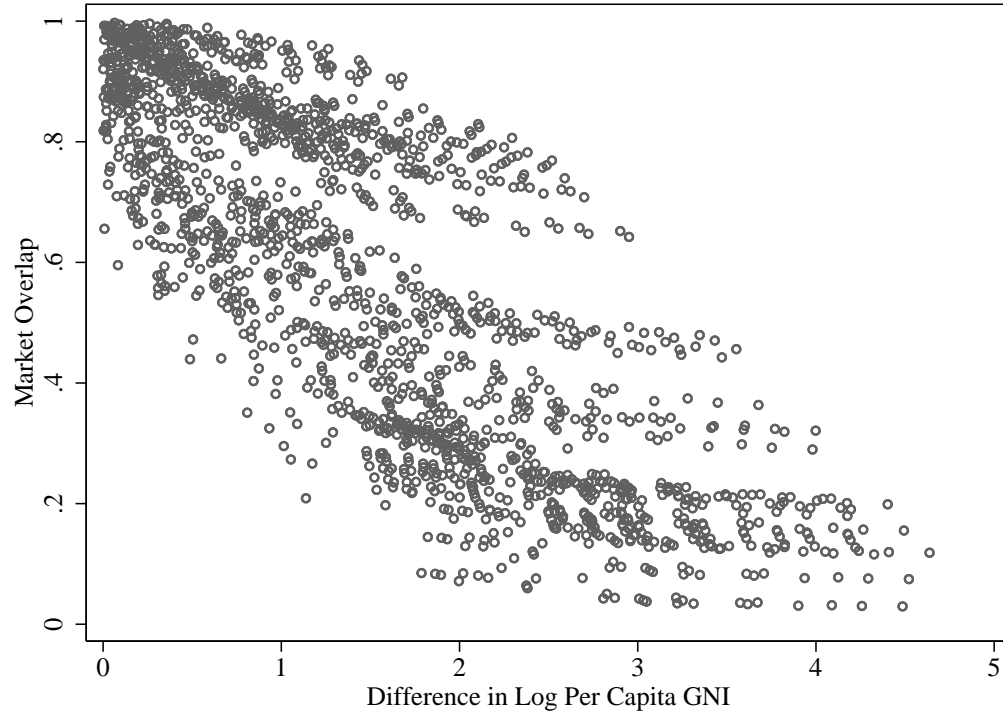**Figure 2.2**: *Market Overlap and Income Differences*
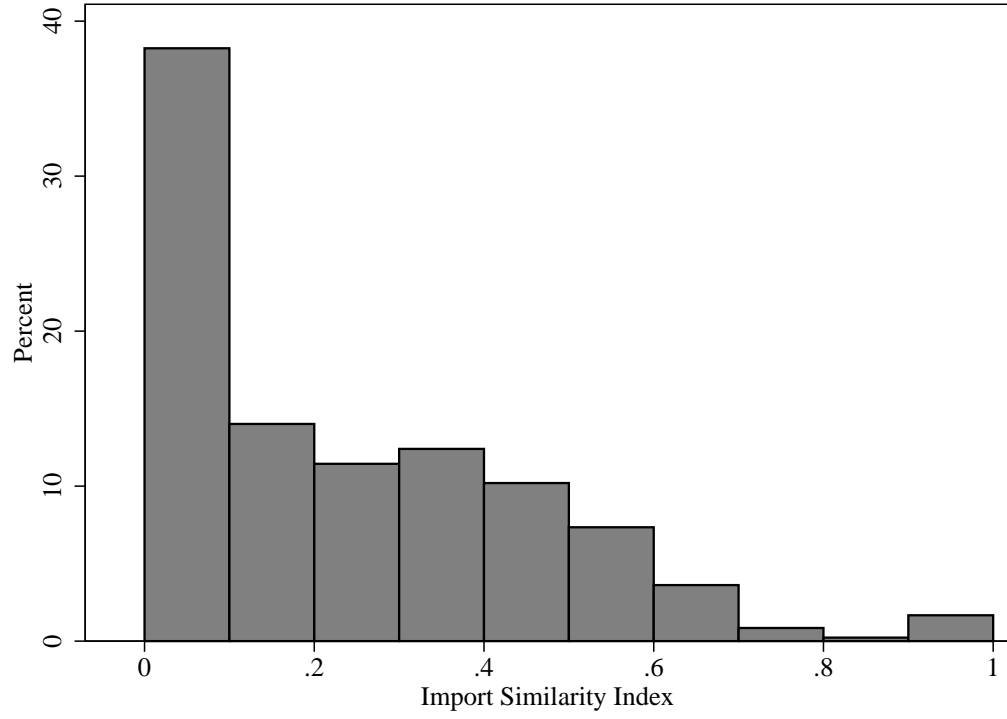
Figure 2.3: *Import Similarity Index*

Figure 2.4: *Import Similarity Index of High Income Countries*
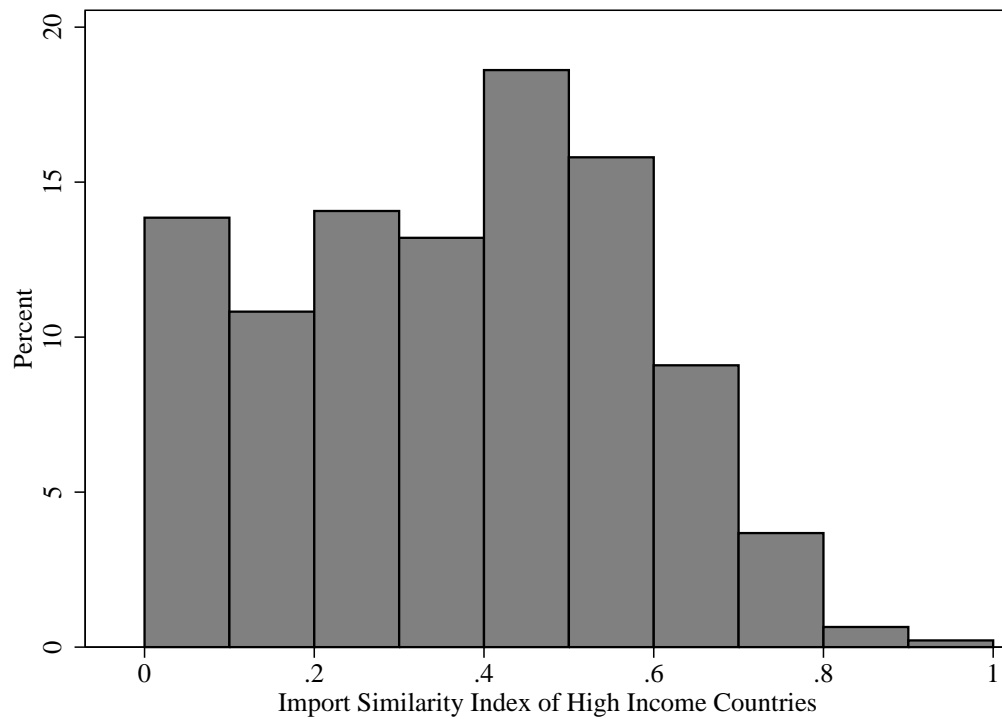
Figure 2.5: *Import Similarity Index of Middle Income Countries*
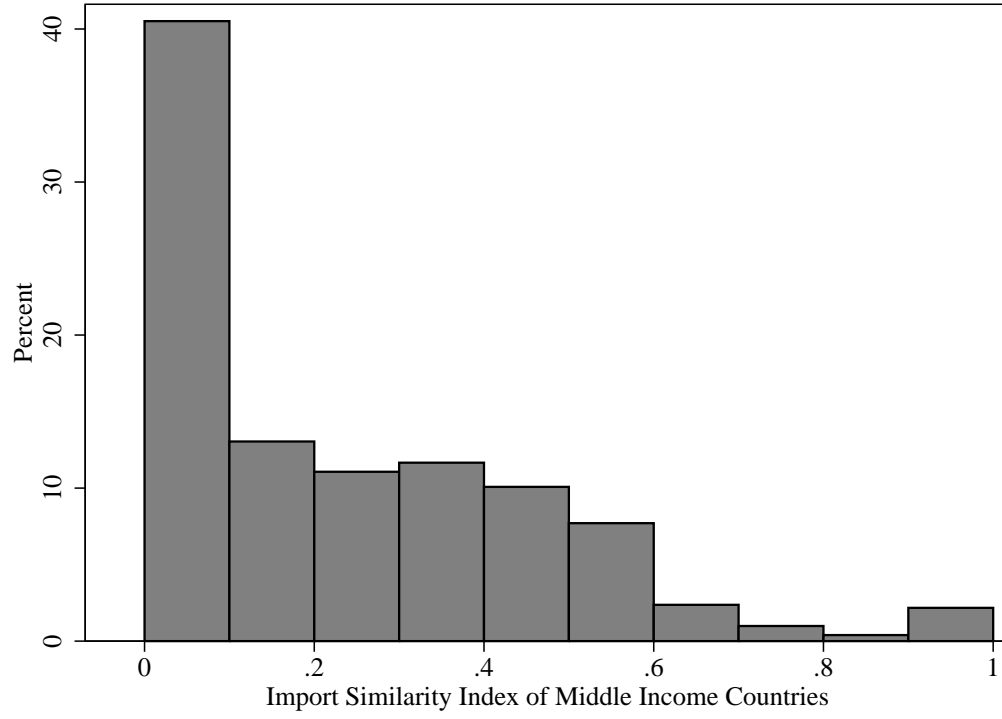
Figure 2.6: *Import Similarity of Low Income Countries*
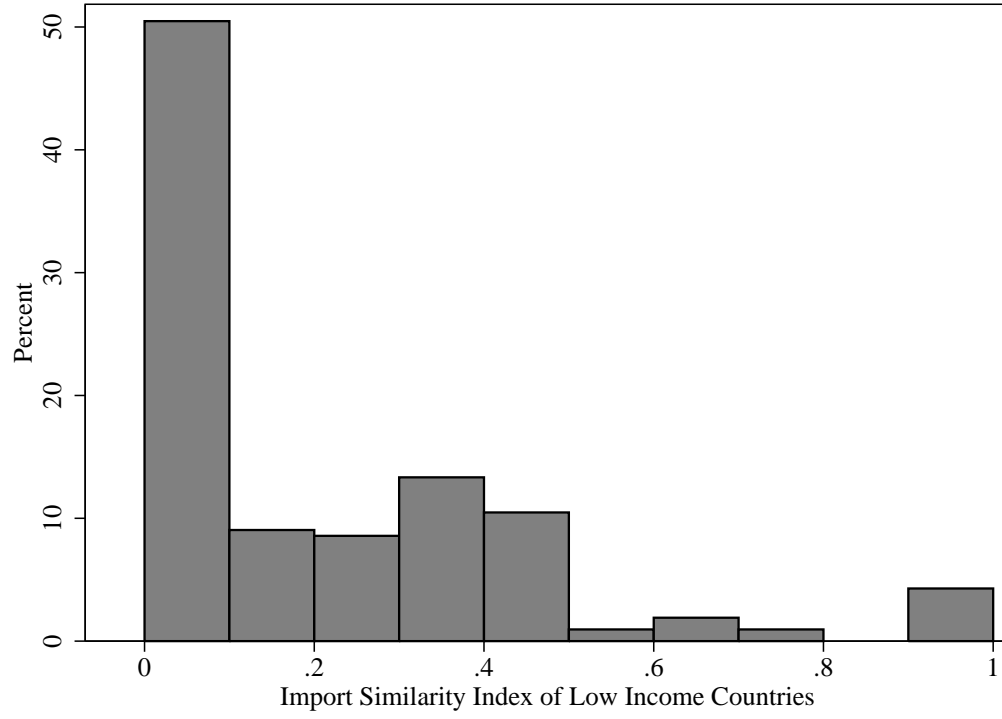
Table 2.2: *Import Similarity*

VARIABLES	(1) ISI	(2) ISI	(3) ISI	(4) ISI
MO	0.163***	0.156***		0.122***
	(0.00930)	(0.00904)		(0.0156)
gnidiff			-0.0389***	-0.00924**
			(0.00231)	(0.00381)
rtadiff		-0.00916	-0.00912	-0.00884
		(0.00587)	(0.00600)	(0.00588)
contigdiff		0.0229**	0.0230**	0.0228**
		(0.00878)	(0.00875)	(0.00877)
comlangdiff		0.00236	0.00142	0.00215
		(0.00704)	(0.00707)	(0.00707)
col45diff		-0.00718	-0.00706	-0.00718
		(0.0133)	(0.0132)	(0.0133)
comcurdiff		0.0236***	0.0199***	0.0230***
		(0.00638)	(0.00631)	(0.00641)
distdiff		-0.0286***	-0.0284***	-0.0286***
		(0.00331)	(0.00330)	(0.00331)
popdiff		-0.00775***	-0.00626***	-0.00739***
		(0.000977)	(0.000979)	(0.00100)
Exporter Effects	Yes	Yes	Yes	Yes
Importer Effects	Yes	Yes	Yes	Yes
Observations	102,660	102,660	102,660	102,660
R-squared	0.385	0.396	0.395	0.396

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 2.3: *Import Similarity*

VARIABLES	(1) ISI	(2) ISI	(3) ISI	(4) ISI
MO	0.134*** (0.00666)	0.126*** (0.00643)		0.0945*** (0.0126)
gnidiff			-0.0318*** (0.00166)	-0.00867*** (0.00320)
rtadiff		-0.00583 (0.00410)	-0.00569 (0.00411)	-0.00551 (0.00410)
contigdiff		0.0120* (0.00719)	0.0120* (0.00718)	0.0120 (0.00719)
comlangdiff		0.00794 (0.00670)	0.00717 (0.00672)	0.00774 (0.00671)
col45diff		0.0143 (0.0103)	0.0143 (0.0102)	0.0142 (0.0103)
comcurdiff		0.00518 (0.00548)	0.00217 (0.00548)	0.00464 (0.00553)
distdiff		-0.0292*** (0.00280)	-0.0290*** (0.00279)	-0.0291*** (0.00279)
popdiff		-0.00623*** (0.000815)	-0.00502*** (0.000815)	-0.00590*** (0.000796)
Exporter Effects	Yes	Yes	Yes	Yes
Importer Effects	Yes	Yes	Yes	Yes
Observations	101,382	101,382	101,382	101,382
R-squared	0.550	0.562	0.561	0.562

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 2.4: *Income Subsamples*

	VARIABLES	(1) ISI	(2) ISI	(3) ISI	(4) ISI
High Income	MO	0.116***	0.116***		0.135***
		(0.0166)	(0.0177)		(0.0345)
	gnidiff			-0.0516***	0.0107
				(0.00817)	(0.0154)
	Additional Controls	No	Yes	Yes	Yes
	Exporter Effects	Yes	Yes	Yes	Yes
	Importer Effects	Yes	Yes	Yes	Yes
	Countries	22	22	22	22
	Observations	4,620	4,620	4,620	4,620
	R-squared	0.603	0.635	0.634	0.635
Middle Income	MO	0.241***	0.196***		0.370***
		(0.0442)	(0.0399)		(0.0775)
	gnidiff			-0.0220**	0.0435**
				(0.00909)	(0.0177)
	Additional Controls	No	Yes	Yes	Yes
	Exporter Effects	Yes	Yes	Yes	Yes
	Importer Effects	Yes	Yes	Yes	Yes
	Countries	23	23	23	23
	Observations	5,313	5,313	5,313	5,313
	R-squared	0.078	0.138	0.137	0.139
Low Income	MO	-0.534**	-0.523**		-1.521**
		(0.204)	(0.188)		(0.566)
	gnidiff			0.0123	-0.0651
				(0.0146)	(0.0393)
	Additional Controls	No	Yes	Yes	Yes
	Exporter Effects	Yes	Yes	Yes	Yes
	Importer Effects	Yes	Yes	Yes	Yes
	Countries	15	15	15	15
	Observations	1,365	1,365	1,365	1,365
	R-squared	0.094	0.122	0.121	0.123

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1 Control variables include:$gnidiff_{bc}$ is the absolute difference of gross national income per capita between importing country b and importing country c. $rtadiff_{bc}$ equals to one if only one of the two importing countries have a trade agreement with exporting country, zero otherwise.$contigdiff$ equals to one if only one of the two importing countries locates next to the same exporting country, and zero otherwise. $comlangdiff$ equals to one if only one of the importing countries and the exporting country use the same language, and zero otherwise. $col45diff$ equals to one if only one of the two importing countries has the colonial relationship with the exporting country prior to 1945, or zero otherwise. $comcurdiff$ equals to one if only one of the importing countries and the exporting country use the identical currency in their respective countries, or zero otherwise. $popdiff_{bc}$ controls for the population difference between the two importing countries b and c.

Table 2.5: *Goods Classification Subsamples*

		(1)	(2)	(3)	(4)
	VARIABLES	ISI	ISI	ISI	ISI
Differentiated Goods	MO	0.105***	0.0935***		0.191***
		(0.0119)	(0.0117)		(0.0518)
	gnidiff			-0.00935*	0.0321**
				(0.00483)	(0.0152)
	Additional Controls	No	Yes	Yes	Yes
	Exporter Effects	Yes	Yes	Yes	Yes
	Importer Effects	Yes	Yes	Yes	Yes
	Countries	60	60	60	60
	Observations	102,660	102,660	102,660	102,660
	R-squared	0.279	0.299	0.294	0.302
Reference Goods	MO	0.0510***	0.0355**		0.0725***
		(0.0154)	(0.0153)		(0.0221)
	gnidiff			-0.00756*	0.0101*
				(0.00387)	(0.00525)
	Additional Controls	No	Yes	Yes	Yes
	Exporter Effects	Yes	Yes	Yes	Yes
	Importer Effects	Yes	Yes	Yes	Yes
	Countries	60	60	60	60
	Observations	102,660	102,660	102,660	102,660
	R-squared	0.285	0.302	0.301	0.302
Homogeneous goods	MO	0.0243	0.00952		0.0374
		(0.0148)	(0.0148)		(0.0297)
	gnidiff			-0.00150	0.00761
				(0.00349)	(0.00608)
	Additional Controls	No	Yes	Yes	Yes
	Exporter Effects	Yes	Yes	Yes	Yes
	Importer Effects	Yes	Yes	Yes	Yes
	Countries	60	60	60	60
	Observations	102,660	102,660	102,660	102,660
	R-squared	0.373	0.387	0.387	0.387

Notes: Robust standard errors in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$ Control variables include: $gnidiff_{bc}$ is the absolute difference of gross national income per capita between importing country b and importing country c. $rtadiff_{bc}$ equals to one if only one of the two importing countries have a trade agreement with exporting country, zero otherwise. $contigdiff$ equals to one if only one of the two importing countries locates next to the same exporting country, and zero otherwise. $comlangdiff$ equals to one if only one of the importing countries and the exporting country use the same language, and zero otherwise. $col45diff$ equals to one if only one of the two importing countries has the colonial relationship with the exporting country prior to 1945, or zero otherwise. $comcurdiff$ equals to one if only one of the importing countries and the exporting country use the identical currency in their respective countries, or zero otherwise. $popdiff_{bc}$ controls for the population difference between the two importing countries b and c.

Table 2.6: *Good Classifications of Imports by Income Level*

	Differentiated Goods	Reference Goods	Homogeneous Goods
High Income Country	69.9%	18.6%	11.5%
Middle Income Country	52.3%	23.6%	24.2%
Low Income Country	37.6%	30.7%	31.7%

Chapter 3

Impulse Response Function Analysis and Empirical Models of Exchange Rate Dynamics

3.1 Motivation

A fundamental question for the economics profession is explaining movements in exchange rates. The exchange rate affects many important variables, including exports, imports, and international capital flows. The focus of this essay is on measurement of the effects of monetary policy on the exchange rate by means of impulse response analysis. Impulse response analysis is a popular tool in time series econometrics that is used to estimate the dynamic response of variables to shocks (Hamilton (1994)). We present monte carlo evidence that the conventional approach to the estimation of vector autoregressive (VAR) models leads to two types of bias in the estimated impulse response functions.[1] Suggestions are provided for dealing with the bias.

The first source of bias follows from the presence of "hump shaped" responses to shocks in the international macroeconomics literature. The best known theoretical example is probably the Dornbusch overshooting model. Dornbusch (1976) presented a model of how changes in monetary policy feed through to cause changes in the exchange rate. An increase in the nominal quantity of money causes either inflation of goods prices or a depreciation of

---

[1]The meaning of "bias" will be explained precisely below.

the exchange rate in the long run. His surprising finding was that the exchange rate depreciates more, relative to the pre-shock exchange rate, in the short run than in the long run. A graph of this "overshooting" of the exchange rate can be seen in Figure 3.1. Dornbusch's model has become a central building block for theoretical models of the exchange rate.

On the empirical side, VAR models and impulse response function analysis have become the main tools for estimating the response of the exchange rate to monetary policy shocks. These tools are used by the researcher as follows. First, she assumes the exchange rate interacts with other variables under the framework of VAR model. For example, letting N be the lag lengths, a VAR model can be written in the following form:

$$M_t = \sum_{i=1}^{N} \beta_{1i} M_{t-i} + \sum_{i=1}^{N} \theta_{1i} E_{t-i} + \gamma_{11} \varepsilon_{M,t} + \gamma_{12} \varepsilon_{E,t}$$

$$E_t = \sum_{i=1}^{N} \beta_{2i} M_{t-i} + \sum_{i=1}^{N} \theta_{2i} E_{t-i} + \gamma_{21} \varepsilon_{M,t} + \gamma_{22} \varepsilon_{E,t} \tag{3.1}$$

E_t is the exchange rate in period t. M_t is a variable that represents monetary policy at period t, such as the federal funds rate. An important feature of the VAR model above is that rather than having an error term, each equation responds to two unobserved structural shocks in the system. The identification problem in impulse response function analysis is to impose enough assumptions on the system above so that all the coefficients, including the γ_{ij}, can be estimated. The impulse response function of the exchange rate following a monetary policy shock, $\frac{\partial E_{t+h}}{\partial \varepsilon_{M,t}}$, can be calculated by simulating on the system when $\varepsilon_{M,t} = 1$ and $\varepsilon_{M,t} = 0$.

As examples of identification, Sims (1992) and Eichenbaum and Evans (1995) assume that monetary policy reacts to a contemporaneous exchange rate shock and its own shocks, but is independent of other variables, while the exchange rate only reacts to its own shock contemporaneously. In contrast, Favero and Marcellino (2004) and others assume the opposite, namely, a contemporaneous shock to monetary policy affects the current exchange

rate, but not the other way around.

Impulse response analysis has found evidence of a hump-shaped response of the exchange rate to a monetary policy shock. For instance, Clarida and Gali (1994) and Eichenbaum and Evans (1995) build VAR models that encompass the theoretical framework of Dornbusch (1976). They find that the peak response of the exchange rate under a contractionary monetary policy does not occur instantly, as predicted by Dornbusch's model. In fact, the peak usually occurs one to three years after the initial monetary policy shock. This phenomenon is known as *delayed overshooting* in the international macroeconomics literature[2]. Faust and Rogers (2003) argue that *delayed overshooting* results from unrealistic assumptions. For example, in Eichenbaum and Evans (1995), the exchange rate did not respond to Federal Reserve policy until a month later (consistent with $\gamma_{21} = 0$). This assumption is dubious at best, since the adjustment of the exchange rate can happen in a matter of seconds. They conclude that the empirical result of *delayed overshooting* is quite sensitive to assumptions. By changing the assumptions, *delayed overshooting* may or may not occur. In another paper, Bjournland (2009) argues that *delayed overshooting* occurs in models that assume no contemporaneous interaction between exchange rate movements and monetary policy. By allowing these two variables to interact contemporaneously, (i.e. γ_s are non zero) *delayed overshooting* disappears.

An important question that arises in light of these empirical results is the following. How well can a VAR model capture a (nonsmooth and non-monotonic) hump-shaped impulse response function? We investigate that question with a series of monte carlo experiments. Using data generating processes (DGPs) motivated by published empirical results, we conclude that the answer to that question is in most cases "not very well". In cases relevant to the study of exchange rates, the impulse response functions estimated in the conventional way exhibit an extreme bias toward zero. Having documented this bias, we investigate the

---

[2]See Cushman and Zha (1997).

performance of two solutions. It is sometimes possible to remove almost all of the bias.

In the second part of the essay, we investigate a related problem with the use of impulse response function analysis for monetary policy. We argue that, in general, a central bank's objective function will not be symmetric in exchange rate forecast errors. Overpredictions and underpredictions of the exchange rate of the same magnitude will not be viewed as equal by the central bank. We suggest a specific loss function for the central bank that incorporates all of the relevant concerns. Simulations show that when the estimation of the underlying VAR model properly accounts for the central bank's loss function, the impulse response function estimates can be very different. We finish by revisiting the paper by Eichenbaum and Evans (1995). We find that their results change when allowance is made for the issues raised in this essay.

3.2   The Bias in Conventional Impulse Response Function Estimates

In this section, we document the bias in conventional VAR model estimates of impulse response functions, when the underlying DGP implies hump-shaped impulse response functions. We proceed as follows. First, we generate 1000 simulated datasets for various DGPs motivated by the empirical literature. Second, a VAR model is estimated for each data set, with the lag length chosen by the AIC, which is the most common method for estimating VAR models (Akaike (1974)). Third, impulse response functions up to 24-steps ahead are calculated for each set of estimates. We compare the average of the estimated impulse response functions at each horizon to the true impulse response function.

3.2.1 Simulation Design and Results

The DGP is a vector moving average (VMA) process with twenty five lags. The DGP can be represented by the equation:

$$X_t = \sum_{i=0}^{24} A_i \epsilon_{t-i} \tag{3.2}$$

where X_t is a vector of two endogenuous variables observed at time t, $X_t = (x_t, y_t)$, T is the number of observations used in this simulation, and A_K is a (2×2) coefficient matrix.

The results of a monte carlo simulation always depend on the choice of parameters, so we provide here some motivation for our choice of parameters. The sample size, T, is 100, 250, or 500 observations. A sample of size 500 corresponds to approximately forty years of monthly data, which at the time of this writing is close to the amount of data available for the post-Bretton-Woods period, which collapsed in 1973. $T = 250$ corresponds to papers that use approximately twenty years of monthly data.[3] Finally, $T = 100$ corresponds to twenty five years of quarterly data.[4]

Since the DGP is stated in moving average form, the elements in the A_K matrix are the true impulse response functions at period k. Once the true impulse response reaches its peak, it gradually declines to its pre-shock average by period 24. The choice of matrices A is based on the behavior of impulse response functions observed in the literature,[5] but we also consider perturbations of the A_K matrix in order to show that the simulation results apply to more than just a single set of potential DGP's. The shapes of the DGPs we consider can

---

[3] As two examples, Eichenbaum and Evans (1995) used 17 years of monthly data, and Faust and Rogers (2003) used 23 years of monthly data.

[4] Clarida, et al (2000) used 30 years of quarterly data in their study of monetary policy rules.

[5] Kim and Roubini (2000), page 573, figure 1, or Bjornland (2009), page 68, figure 1.

be seen in Figure 3.2.[6] The true impulse response reaches its peak in the third, sixth, tenth or fifteenth period after the initial shock. This assumption corresponds to the results in the literature that the shape of the impulse response following a monetary policy shock peaks after one to three years. If we were to include longer peak times, the bias would be even greater.

For each combination of T and A_k, we generate 1000 samples of data assuming the errors ϵ are independent standard normal random variables, estimate a reduced form VAR with the lag length selected by the AIC, and apply a Choleski decomposition to the reduced form residual covariance matrix to recover the impulse response functions. We then compute the mean of the impulse response function across the 1000 simulated datasets. The primary focus is on the difference between (i) the number of periods after the shock that the true impulse response function peaks and the number of periods after the shock that the average estimated impulse response function peaks, and (ii) the difference between the magnitude of the peak for the true and average estimated impulse response function.

Figure 3.3 plots the average impulse response function (IRF) of variable y following a one unit shock to x for sample size $T = 100$. Each figure corresponds to one of the four peak periods, with the true impulse response function in blue and the average estimated impulse response function in red. The conventional impulse response function is accurate when the IRF peaks in the third period. The hump shape of the true IRF does not have much effect so long as the peak occurs early. Moving on to the other plots, we see that the quality of the IRF estimates deteriorates rapidly when the IRF peaks at a later date. When the peak occurs in the 15th period after the shock, the average estimated IRF peaks at about 20% of the true value, and the estimated peak occurs much earlier than the true peak. For a small sample with 100 observations, the conventional VAR estimation strategy leads to disastrous resulsts. The AIC selects a model that is not capable of accurately capturing the dynamics

---

[6]Figure 3.2 does not show all the DGPs, only the different shapes that we considered. For each shape, we varied the sample size and the estimation methodology.

of the system. In the next section, we provide evidence that the problem is that the usual method of selecting lag length chooses models that are overly restricted, causing them to fail to properly capture the nonmonotonic behavior of the impulse response functions.

It is easy to observe from these figures that OLS estimation of VAR models with AIC lag selection does not deliver anything close to unbiased impulse response function estimates when the true impulse response functions are hump-shaped. Moreover, as the timing of the peak period moves further away from the initial shock, the bias of conventional impulse response function estimates gets even worse.

We report the size of the bias at the peaking period in Table 3.1. The first column of Table 3.1 represents the size of the bias of the OLS based impulse response functions with different true impulse response functions. The second column reports the period in which the estimated impulse response peak. For example, with $T = 100$, if the true impulse response peaks at three period after an initial shock, the estimated impulse response suffer 9 % bias as it peaks at five period after an initial shock. These results in this table is consistent with the impulse response functions we observed in Figure 3.3. The size of the bias increases as the true impulse response funciton peaks at the longer horizon.

Figure 3.4 and 3.5 report the OLS based impulse response functions with $T =$250 and 500, respectively. Judging from these figures, the impulse response functions estimated using AIC lag selection still perform poorly when the true impulse response functions peak in a later period. The size of the bias can be found in the third column and the fifth column of Table 3.1 with $T =$250 and 500, respectively. While the size of the bias declines as more observations are avaliable for estimation, the improvements are not significant. For example, when the true impulse response function peaks in the sixth period, doubling the number of observations from 250 to 500 decreases the size of the bias by only 0.8%. Clearly, increasing the number of observations does not eliminate the bias.

Overall, these results suggest that the conventional method for selecting and estimating

70

a VAR model is the source of the incredible bias when the true impulse response functions are hump-shaped. It is plausible that the source of the problem is that the AIC favors small models. That may be fine when the goal is to produce a one-step ahead forecast, but it clearly does not work very well when the researcher is estimating impulse response functions many steps ahead, if the true impulse response function is not monotonically decreasing.

3.2.2 Two Approaches to Reduce the Bias

This section introduces two possible solutions to the problem of bias in impulse response function estimation, increasing the lag length of the VAR model, and changing the loss function used to estimate the VAR model. We then provide simulation evidence on the ability of both solutions to reduce or eliminate the bias of estimated impulse response functions. We find that a sufficiently long lag length will almost completely eliminate the bias. The drawback to that approach is that in practice the researcher only has rough guidance about the choice of lag length for the VAR model. On the other hand, the use of a loss function based estimator allows the researcher to use existing model selection methods, but does not completely eliminate the problem of bias.

3.2.3 Solution 1: Increasing the Lag Length

A VAR model is a linear model, so if the world is nonlinear, it should be viewed as nothing more than a linear approximation. The choice of lag length for the estimated model affects the accuracy of the approximation. Thus, lag length selection is essential to the estimation of a VAR model for purposes of impulse response construction. When the underlying DGP exhibits a large degree of nonlinearity, it is necessary to have a sufficiently long lag length to capture all of the nonlinear behavior. Thus, the number of lags required is correlated

with the degree of nonlinearity in the system.

It is possible that the poor performance of conventional VAR models is due to the choice of an excessively parsimonious model. The AIC may be poorly suited to the task of selecting lag length in models with hump shaped impulse response functions. Capturing hump shaped impulse response functions may require many more lags than the AIC selects to minimize one-step ahead forecast errors.

We investigate the possibility that the AIC is the cause of the bias by doing a new set of monte carlo experiments. In this exercise, we continue to assume that the underlying DGP follows equation 3.2. We also assume that the true impulse response functions peak in the sixth or fifteenth period after an initial shock. We restrict the size of the observation to 250, i.e. $T = 250$. The difference in this set of simulations is that we fix the number of lags used to estimate the VAR models to some predetermined level. Although such knowledge is rarely available in practice, a finding that the bias disappears completely or mostly when fixing the lag length to a predetermined level would tell us at least two things:

(i) the source of the problem is that the AIC selects an insufficient number of lags when the true impulse response functions are hump shaped, and

(ii) the AIC provides only a lower bound on the number of lags that should be used, and in practice the researcher should check the robustness of the estimated impulse response functions to the inclusion of more lags.

We fix the lag length of the estimated VAR model to three, six, nine, twelve, fifteen and eighteen lags in each of the simulations.

3.2.4   Solution 2: Loss Function Based Estimation

The second solution that we propose is to use a loss function based (LFB) estimator. An LFB estimator minimizes the loss function that accounts for the loss of the decision maker.

Specifically, we propose to estimate the reduced form VAR model by solving the following optimization problem:

$$L = min \frac{1}{T} \sum_{t=1}^{T} (\widehat{x}_{t+h} - x_{t+h})^2 \qquad (3.3)$$

where T is the number of observations and h is the time horizon of the impulse response function. Thus, the LFB estimation approach requires estimation of the parameters of the VAR model for each time horizon at which an IRF is calculated. If impulse response functions are calculated for horizons 1 through 24, the parameters of the VAR model will have to be estimated 24 times.

By directly accounting for the fact that the goal is to make an h-step ahead forecast, it is expected that the nonmonotonic and nonlinear features of the IRF will cause fewer problems. A VAR model estimated by OLS is the special case of minimizing the above loss function for $h = 1$ no matter the horizon. If the goal is only to get consistent estimates of the parameters of the VAR model, OLS is a reasonable strategy. Unfortunately that is not what we are trying to do. Therefore it may help to minimize the relevant loss function.

3.2.5 Simulation Results: Longer Lag Lengths

In this section, we show the simulation results of the first solution we proposed, that is, we fix the lag length of the estimated VAR model to a predetermined value. Figure 3.6 and 3.7 plot the OLS based impulse response functions of variable y to a one unit shock to x with $T = 250$, when the peak is in period six and fifteen, respectively. Each figure represents a different lag length (3, 6, 9, 12, 15, or 18). The true impulse response function is represented by the blue line in each graph. The red line represents the average of the impulse response functions over the 1000 simulated datasets.

Several key results emerge from these figure: First, in comparsion to the performance of the VAR models with AIC lag selections, the performances improves when more lags are used

in the estimations. The gap between the true impulse response functions and those obtained from the VAR become smaller, especially at the short and medium horizon. Second, when the number of lags used in the estimations is the same as the number of periods it takes for the true impulse response functions to peak, the size of the bias at the peaking period is at the minimum. Table 3.2 report the size of the bias at the peaking period. The first and second column of Table 3.2 represent the sizes of bias in which the true impulse response peaks at sixth period and fifteen period after the initial shock, respectively. Each row of the same table represents a number of lag length used in estimations. From the first column of Table 3.2, we observe that the size of the bias is at the minimum when six lags are used in the estimations. Similarly, when the true impulse response functions peak at fifteen period after the initial shock, the minimum bias at peaking period happens in the estimations with fifteen lags, 0.107, from the fifth row of the second column in Table 3.2. Third, from the same table, we observe that when the number of lags used in the estimations is further away from the number of period it takes for the true impulse response to peak, the size of the bias at the peaking period increases. For example, in the first column we observe that when fifteen lags are used in the estimations, the size of the bias is aproximately 0.087. In comparison, when eightenn lags are used in the estimations, the size of the bias increases to 0.103. Fourth, when more lags are used in the estimations, the estimated impulse response capture the shape of the true impulse response better.

These results suggest that if the researcher knows exactly when the true impulse resposne functions peak, he should use the number of lags that is equivalent to the amount of times it takes for the true impulse response to peak. For example, if the researcher knows the true impulse response function peaks at sixth period after an initial, he should assign six lags in the estimations. The problem about this solution is that usually the true researcher has no prior knowledge about when the true impulse response functions peak, therefore, he does not know the optimal number of lags that he should use in VAR models.

74

3.2.6 Simulation Results: Loss Function Based Estimation

In this section, we report the results of the second solution in Figure 3.8 and 3.9. In these figures, we again plot the impulse response functions of variable y to a unit shock of x with $T = 250$ and with two different DGP. In these figures, the true impulse response functions are represented by the blue line. The red and green lines represent the impulse response functions obtained from OLS and LFB estimations, respectively.

A number of key results emerge from these figures: First, when more lags are used in LFB estimation, the performance improves. In Figure 3.8, when three lags are used in the estimation, the LFB impulse response functions diverge dramatically after the peaking period. More specifically, it does not converge to zero at long horizons. Second, in most cases, the size of the bias at the peak is smaller in LFB estimation than in OLS estimation. Table 3.3 report the size of the bias with LFB estimation at the peaking period. The first and third column of this table report the bias where the true impulse response peak at sixth and fifteen period, the second and the fourth column report the period in which the estimated impulse response peak respectively. In comparison to Table 3.2, in a majority of cases, the size of the bias under LFB estimation is smaller than those obtained from the OLS. Exceptions occur in the case in which three lags are used in the LFB estimation, or in which the true impulse response functions peak in the sixth period and at the same time, six lags are used in the OLS estimations. Third, it appears there exists an optimal number of lags that should be used in LFB estimation. For example, from the first column of Table 3.3, we observe that bias is at a minimum when nine lags are used in the estimation. Fourth, at short and medium horizons, when more lags are used in estimation, the LFB impulse responses stay closer to the true impulse response function than those obtained from the OLS. Finally, we observe when more lags are used in estimation, the estimated impulse responses obtained from the LFB capture the shape of the true impulse response functions better.

Several lessons can be learned from this series of simulations. First, if the researcher believes the true impulse response function is hump-shaped (something that can be observed by looking at the impulse response function given by OLS), the AIC is not an appropriate criteria for lag length selection. A minimum response that should be taken is to see if the results change when using more lags than the AIC recommends. Second, if the researcher has prior information about when the true impulse response peaks, he should choose the number of lags that is equivalent to the amount of time for the true impulse response functions to peak in the VAR model. Third, in most cases, the LFB estimator of the impulse response function is preferable to those obtained from OLS estimation, since the size of the bias is typically smaller with LFB estimation.

Perhaps the most important lesson from these results is that we should be skeptical of the empirical evidence in the international macroeconomics literature that has relied on VAR models. In the international macroeconomics literature, hump-shaped impulse response functions are common. In addition, quite often the researcher does not know how the true impulse response behaves. With these things in mind, the simulation results we have presented suggest that the researcher should definitely not be content to draw conclusions from a single model specification.

3.3 Optimal Exchange Rate Forecasts for Monetary Policy

It is not uncommon for monetary authorities to follow a policy rule such as the "Taylor rule"(Taylor (1993)) to set a target for the interest rate(Ball (1997), Orphanides (2002 2003)). In the process of pushing the interest rate to its target level, the monetary authority causes changes in the money supply and in turn, the exchange rate(Clarida et al. (1998 2000 2002)). It is natural that a central bank will be concerned about the level of the exchange

rate. The level of the exchange rate might even be one of the tools used in the monetary policy process(Svensson (2003 2004)).

Even if the exchange rate does not play a direct role in monetary policy, there are obvious reasons for a central bank to consider the impact of its actions on the exchange rate. A favorable exchange rate can stimulate economic growth (Levy-Yeyati and Sturzenegger (2005), Rodrik (2008)). A depreciation of the local currency increases the competitiveness of domestic exporters. On the other hand, depreciation of the domestic currency is the equivalent of an appreciation of foreign currencies, causing foreign goods to become more expensive, which increases the cost of living.

These issues have been at the heart of many recent news stories about the People's Republic of China[7]. The U.S. and other countries have criticized the Chinese government for following a policy that keeps the renminbi "undervalued" in an attempt to provide Chinese manufacturers an unfair advantage in the markets in which their exports compete. As of now, it remains unclear how the U.S. and Chinese given can resolve their disagreements in this area.

It is not necessarily the case that the central bank will prefer to keep the currency at a low level. The Asian financial crisis in 1997 showed how important it is for central banks to be able to control the value of their currencies[8]. In 1997, a series of financial attacks by international speculators lead to a series of depreciations of Asian currencies. These depreciations created economic turmoil across the region. For example, in the case of Thailand, the Thai baht depreciated over forty percent between June 1997 and July 1998. The CIA workbook[9] reports that Thailand GDP per capita decreased from $8,800 in 1997 to $8,300 in 2005. Moreover, these crises caused political instability. In Thailand, it led to the resignation of Prime Minister General Chavalit Yongchaiyudh[10]. These political crises

---

[7]See?? (BBC), Krugman (2010), McDonald (2011), McQuillen and Tracer (2011)

[8]See Hunter et al. (1999)

[9]See Yellen (2007)

[10]See Kate and Onsanit (2008)

further handcuffed the government's ability to react.

Central banks need to concern themselves with not only the level of the exchange rate, but the volatility as well. Volatility of the exchange rate adds uncertainty for households and firms participating in international goods and financial markets. Many papers have looked at exchange rate volatility, in particular using GARCH models (Alexander and Lazar (2006), Kearney and Patton (2000), Patton (2006), Vilasuso (2002)). Reduction of uncertainty about the exchange rate is one of the reasons that floating exchange rates are a relatively recent phenomenon. Before the floating exchange rate regime, the exchange rate was determined by the Bretton-Woods agreement[11] Under this regime, the value of foreign currency were pegged to the value of the U.S. dollar. In addition, each member is required to maintain exchange rates within plus or minus 1% of parity by controlling the amount of foreign currency in the international financial market. In the case of the U.S. dollar, the U.S. government agreed to link the value of the dollar to gold at a rate of $35 dollar per ounce. Foreign governments and central banks were allowed to exchange dollars for gold at this rate.

This fixed exchange rate regime lasted until 1967 when the value of the UK pound was under attack. The UK government was forced to devalue the sterling more than the Bretton-Woods agreement allowed[12]. This devaluation put a tremendous pressure on the value of the U.S. dollar and the quantity of gold reserves. More specifically, the uncertainty about the amount of gold reserves required to maintain a fixed exchange rate regimes lead to a failure of this system, as President Johnson states[13], "The world supply of gold is insufficient to make the present system workable—particularly as the use of the dollar as a reserve currency is essential to create the required international liquidity to sustain world trade and growth. This crisis lead to the present flexible exchange rate regime in international financial markets. All else equal, central banks would like to reduce uncertainty about the

---

[11]See Dormael (1978)

[12]See BBC (BBC)

[13]Gavin (2003)

exchange rate to zero.

In what follows, we show how the arguments in the preceding paragraphs have direct implications for the loss function of the central bank. We motivate a loss function that reflects the central bank's concerns. We then show that the impulse response functions one estimates using the conventional approach can sometimes be very different from impulse response functions that are estimated taking into account the central bank's loss function.

3.3.1 The Central Bank's Loss Function

The discussion above indicates that one of the goals of a central bank is to make a decision about the desired level of the exchange rate. Even if the central bank does not directly target the exchange rate, the variables that are the goals of monetary policy, such as inflation and unemployment, are unlikely to be determined independently of the exchange rate.[14]

Although most central banks probably have a desired level of the exchange rate, they are unlikely to carry out monetary policy in an attempt to immediately push the exchange rate to its desired level. Theoretical arguments have been made for gradual adjustment of monetary policy to its target (Rudebusch (2002)). Empirical work has shown that central banks act gradually, reflecting a preference for small, repeated changes in policy as opposed to large, one-time changes (Sacka and Wieland (2000)).

Consider the following scenario. Assume that the current exchange rate is 1.0 US\$/UK pound and the target exchange rate is 2.0. The optimal monetary policy[15] will move the exchange rate to a value less than 2.0, such as 1.5. In order to determine the optimal federal funds rate target, the Federal Reserve needs an estimate of the effect of monetary policy on the exchange rate, and this information is given by estimated impulse response functions.

---

[14]The Full Employment and Balanced Growth Act (also known as the Humphrey-Hawkins Act) mandates that monetary policy should strive for four objectives: full employment, growth in production, price stability, and balance of trade and budget. In response to this law, one of the objectives of the the Federal Reserve Bank has been to promote "effectively the goals of maximum employment, stable prices, and moderate long term interest rates. See Steelman (2011) for more details.

[15]As stated above, the Federal Reserve may be targeting variables such as the inflation rate, but the optimal monetary policy will still take account of the exchange rate.

If the impulse response functions are estimated by OLS, the loss function is assumed to be quadratic in the forecast errors. Importantly, the use of OLS implies symmetry. Forecasts of 1.25 or 1.75 are equally bad when the target is 1.50, because they are both 0.25 units from the true value.

When a central bank acts gradually, quadratic preferences are not consistent with the central bank's preferences. A forecast of 1.25 is much worse than a forecast of 1.75 because it is farther from the desired exchange rate level. It is not only the deviation from the short-run target that matters, because the central bank also cares about the deviation from the desired (long-run) exchange rate. The magnitude as well as the sign of the forecast error is important.

We now propose a loss function that incorporates these features. The arguments to the loss function need to include the current exchange rate, the desired (long-run) exchange rate, the forecast error, and a penalty variable that depends on the sign of the forecast error. One such loss function is:

$$L = \sum_{t=1}^{T} (\phi_t \hat{\varepsilon}_t)^2 \tag{3.4}$$

where

$$\phi_t = \begin{cases} exp(\alpha(E_c - E^t)) & if \hat{\varepsilon}_t > 0 \\ 1/exp(\alpha(E_c - E^t)) & if \hat{\varepsilon}_t < 0 \end{cases}$$

In equation (3.4), $\hat{\varepsilon}_t = E_t - \hat{E}_t$ is the forecast error of the exchange rate E_t at period t, E_c is the current exchange rate, E^t is the target annual inflation rate, and α determines the amount of asymmetry. The greater α is, the greater the asymmetry is. If the current exchange rate equals the target exchange rate, or if $\alpha = 0$, ϕ_t is equal to one for all observations and equation (3.4) reduces to a quadratic loss function.

3.3.2 Simulation Design

The DGP for this set of simulations is a VAR model estimated using monthly data for the period January 1975 to May 1990, the beginning of the flexible exchange rate regime. In this series of simulation exercises, we want to illustrate the difference between the OLS and the LFB estimates of the impulse response functions. We collected the data on the federal fund rate and the spot exchange rate from the Federal Reserve Bank of St. Louis, FRED database. The spot exchange rate is measured in terms of the US$/UK pound. Each dataset consists of 250 observations, equivalent to approximately twenty years of monthly data. We consider a one unit positive shock to the fed funds rate to be a contractionary monetary policy shock. The federal funds rate does not affect the exchange rate contemporaneously, as in Sims (1992) and Eichenbaum and Evans (1995). The target exchange rate is assumed to be equal to the mean of the log of the exchange rate over the sample period.

There are two objectives in this series of simulations. First, we want to see how the relationship between the current and target exchange rate affects the results. We consider two cases here: the current exchange rate is either above or below the rate the Fed target. We set the current exchange rate to either one standard deviation above or below the target exchange rate (i.e., one standard deviation above or below the mean). Second, we want to identify the role of α, the parameter that determines the degree of asymmetry of the loss function. We consider two cases. We set the degree of asymmetry (α) equal to five initially. With this setting, ϕ , the parameter that penalizes positive and negative forecast errors asymmetrically, is equal to 2.42 or 0.41. In other words, we penalize a typical positive forecast error 5.9 times more than a typical negative error. We also set $\alpha = 0.5$, for which we penalize a typical positive forecast error 1.5 times more than a typical negative error, in a separate series of simulations to analyze how changes in α affect the results.

In each series of simulations, we generate 1000 samples of data assuming the errors ϵ_t are independent standard normal random variables, estimate a reduced form VAR model by

81

OLS, and apply a Choleski decomposition to the reduced form residual covariance matrix to recover the impulse response functions. For the LFB estimator, we assume equation (3.4) is the objective function the policymaker follows. We then compute the mean of the impulse response function across the 1000 simulations.

3.3.3 Simulation Results

We first investigate the case in which the current exchange rate is above the target. Figure 3.10 reports the results in which $\alpha = 5$. Each figure represents a different number of lags used in the estimation. The red and green lines represent the traditional VAR approach (estimated by OLS) and the LFB impulse response functions, respectively.

Several key results emerge from these figures. First, the LFB impulse response functions always lie above those estimated by OLS at short horizons. To understand why this is consistent with our *a priori* expectations, think about the relative effect of an overestimate versus an underestimate of the effect of a monetary policy shock. An overprediction will cause monetary policy to be loose relative to what is necessary to hit the target exchange rate. The money supply will be rising more quickly than optimal, and the exchange rate will depreciate beyond the target. The same logic implies that an underprediction will lead to too little depreciation. Because the exchange rate is above the target, excessive depreciation is not as costly as failing to depreciate. The optimal estimate of the impulse response function should have an upward bias if one directly accounts for the central bank's loss function.

The differences depend on the number of lags used in the estimation. When three lags are used, the LFB estimate predicts a 0.07 percent appreciation of U.S. dollar, whereas the OLS estimate predicts an appreciation of a little bit less than 0.01 percent. On the other hand, when nine lags are used in the estimation, the LFB estimate predicts a 0.04 percent appreciation of U.S. dollar. In contrast, the VAR model estimated by OLS predicts a depreciation of 0.002 percent. Moreover, from Figure 3.10 we observe that when a small

number of lags is used in estimation, the deviations between the OLS and the LFB impulse response functions last shorter, but the magnitude of deviations is large. In contrast, when more lags are used in estimation, the magnitude of deviations is small, but the duration of the deviations is longer.

Next, we want to see how changes of α affect the LFB impulse response functions. We reduce α to 0.5 and plot both LFB impulse response functions in Figure 3.11. First, we observe that when only a few lags are used in LFB estimation, the impulse response with $\alpha= 0.5$ lies above the one estimated with $\alpha= 5$ at the short horizon. Opposite is true if more lags are used in the estimations. For instance, with $\alpha= 0.5$ and three lags are used in estimations, a contractionary monetary policy lead to an appreciation of 0.06 percent initially. In comparison, when $\alpha = 5$, the exchange rate appreciates initially by about 0.048 percents. Second, in all cases, a depreciation of the U.S. dollar happens initially, regardless the degree of asymmetry of the objective functions. Third, at the medium and long horizon, the degree of asymmetry have lesser effects on the impulse response functions. As we observed, impulse response functions fluctuates around each other. Moreover, when more lags used in the estimations, the impulse response functions become more volatile at longer horizon.

We now draw our attentions toward cases in which the current exchange rate is one standard deviation below the target exchange rate. Figure 3.12 reports the results with values of α identical to those in Figure 3.10. Also, the red and green line represents the OLS and LFB impulse response functions, respectively.

We can make similar observations to the case above. Now that the current exchange rate is below the target exchange rate, the LFB estimator impulse response functions lie below those estimated from the VAR models. The same logic described above explains why that is what we expect.

Similarly, we want to investigate the role of α under this circumstance. We assign $\alpha= 0.5$ and plot those impulse response functions in Figure 3.13. Each figure represents a different

number of lags used in the estimation. Similar to Figure 3.11, in each figure the red and green lines represent the LFB impulse response functions with more asymmetry ($\alpha = 5$) and with more symmetry ($\alpha = 0.5$), respectively. Figure 3.13 is just a reflection of Figure 3.11. This is not surprising, since the only difference between these two figures is that the current exchange rate is below the target exchange rate in the first case.

To summarize, there are two lessons we can learn from these exercises: First, as we have learned in previous section 2, it is important to select the number of lags used in the estimations. We observe that different number of lags results in different estimated impulse response functions. Second, the degree of asymmetry in the objective loss functions affects the results dramatically. Different degrees of asymmetry lead to different simulation results.

3.4   Empirical Example

In this section, we provide an example of LFB impulse response function estimation in practice. Specifically, we revisit the paper of Eichenbaum and Evans (EE, 1995) assuming the Federal Reserve's loss function is given by equation (3.4). EE investigate the effect of U.S. monetary policy shocks on the US/UK exchange rate using a VAR model with the recursive ordering $[Y, P, Y^{FOR}, R^{FOR}, FF, NBRX, s_R^{FOR}]$. With this ordering, a shock to a variable x does not have a contemporaneous effect on variables that are to the left of x. For example, a shock to FF exerts contemporaneous effect on itself, $NBRX$ and s_R^{FOR} only. Variables Y and Y^{FOR} are U.S. and UK industrial production, R^{FOR} is the short term UK interest rate, P is the U.S. Consumer Price Level, $NBRX$ represents the ratio of nonborrowed to total reserves, and s_R^{FOR} represents the nominal exchange rate, measuring in US dollar/UK pound. We use a dataset that is almost identical to the one used by EE.[16]

---

[16]There are two differences between the dataset Eichenbaum and Evans used and the dataset we use here. First, Eichenbaum and Evans collects the UK data on industrial production and short term Treasury bill rate from *International Financial Statistic* database. We collect the short term Treasury bill rate from the Bank of England and the UK industrial prodution from The Office for National Statistics. Second, the sample period range from 1974:1-1990:5 in Eichenbaum and Evans. Our sample period range from 1975:1 - 1990:5. The earliest period the Office for National Statistics have on industrial production data is 1975.

All variables are in logarithms except R^{FOR} and FF.

There are two main objectives in this section. First, we want to show any differences between the impulse response estimated using OLS, as in EE, and LFB estimates. Second, we want to see how ϕ, a parameter that represents the amount of punishment imposed on positive and negative forecast error, affects the estimation results.

3.4.1 Empirical Results

Figure 3.14 reports the impulse response function of the exchange rate following a contractionary monetary policy shock, with $\alpha = 5$ and with current exchange rate is one standard deviation above the target exchange rate. In each figure, red and green lines represent the impulse responses obtained from OLS (EE) and the LFB estimator, respectively. We report the impulse response functions for 24 periods.

We observe several key results from Figure 3.14. First, we notice that at short horizons, there is no significant differences between OLS and LFB estimations. At medium and long horizons, the deviations between OLS and LFB estimations depend on the number of lags used. Recall that the objective function (equation (3.4)) depends on two factors: ϕ, a parameter that represents the amount of penalty on positive forecast errors, and ε_t, the residual for period t. If ϕ is small, or if ε is small, or both, the LFB impulse response functions and those estimated from OLS will be more similar. In this seven variables model example, and with this specific dataset, the residuals at the short horizon are relatively small and as a result, the differences between the impulse response obtained from OLS and LFB estimations are not significant.

At the medium and in some cases, long horizon, we observe significant deviations between the LFB impulse response and those estimated from OLS, due to the size of the residuals at these horizons are relatively large. Under this scenario, the punishments on positive(negative) forecast errors become larger(smaller). As a result, the LFB impulse

response functions and those estimated from the OLS are quite different. To verify this claim, we reduce the amount of punishments on positive forecast errors. With this change, we expect the deviations between these two estimated impulse response functions should become smaller.

In Figure 3.15, we plot two estimated impulse response functions. In each figure, the red line represents the OLS impulse response functions and the green line represents the impulse response functions computed from LFB estimation with $\alpha = 0.5$. In these figures, we observe that by reducing the penalty on positive forecast errors, these two estimated impulse response functions are now statistically identical at the medium and long horizons. Therefore, we conclude that the deviations we observe at medium and long horizons in Figure 3.14 are the results of large residuals and large α.

We now turn our attention toward the case in which the current exchange rate is one standard deviation below the target exchange rate. Figure 3.16 reports the results for $\alpha = 5$. In each figure, the red line and green line once again represent the impulse response functions estimated by OLS and the LFB, respectively.

From Figure 3.16, we observe once again that these two estimated impulse response are statistically identical at the short horizons and deviate at the medium and long horizons. We can use the same logic from above to explain why this happens. Figure 3.17 plots the LFB estimation with $\alpha = 0.5$ and we once again observe that by reducing the penalty on positive forecast errors, the deviations between the LFB impulse response functions and thse estimated by the OLS become insignificant at these horizons, regardless the number of lags used in the estimations.

We also notice that the LFB impulse response functions lie below those estimated by OLS in most cases. These results are consistent with what we expect. Recall that in these cases the current exchange rate is below the target exchange rate, so by the same logic we described in section 3, there should be no surprise that the LFB impulse response function lie below than those estimated from the OLS.

3.5 Conclusion

In this chapter, we investigated problems with impulse response analysis, a tool commonly used to study the relationship between monetary policy and exchange rate movements. In the first part of this essay, we investigated the size of the bias of the VAR model under OLS estimation, given the true impulse response function is hump-shaped. We found that VAR models that rely on the AIC for lag selection perform poorly. We suggested two possible solutions. First, increase the number of lags that are used in the estimations, and second, estimate the impulse response functions using a loss function based estimator. We found that both solutions were able to improve the simulation results dramatically. Moreover, these simulation results suggest that we should be cautious when interpreting any international economic literature that include the VAR models estimated by OLS.

In the second part of the essay, we considered a second problem with OLS estimation of impulse response functions. We proposed an asymmetric loss function that the monetary authority could use to estimate the impulse response functions. We then analyzed the difference between the LFB impulse response and those estimated by OLS(with symmetric quadratic objective function). We found that the size of the residuals and punishment of forecast error, determine the differences between LFB impulse response functions and those estimated by OLS.

Table 3.1: *Size of the Bias of the VAR based Impulse Response Functions at the peak period.*

True Peak	T=100	Period	T=250	Period	T=500	Period
3rd Period	-9.0%	5	-4.4%	5	-3.1%	5
6th Period	-41.4%	6	-38.7%	6	-37.9%	6
10th Period	-63.8%	6	-61.9%	6	-61.5%	6
15th Period	-76.3%	7	-75.2%	7	-74.6%	7

Note: The number of lags used in the VAR model is determined by the AIC.

Table 3.2: *Size of the Bias of the VAR based Impulse Response Function at the peak period.*

Number of lags	True IRF peak at 6th period	Period	True IRF peak at 15th period	Period
3	-13.7%	6	-58.9%	9
6	-4.4%	6	-42.2%	11
9	-5.5%	6	-27.4%	13
12	-7.0%	6	-15.4%	14
15	-8.7%	6	-10.7%	15
18	-10.3%	6	-12.9%	15

Note: 250 observations are used in the estimations.

Table 3.3: *Size of the Bias of the loss function based Impulse Response Function at the peak period.*

Number of lags	True IRF peak at 6th period	Period	True IRF peak at 15th period	Period
3	27.3%	8	-30.0%	10
6	-7.4%	6	-26.8%	11
9	-1.2%	6	-17.4%	13
12	-2.1%	6	-10.7%	14
15	-5.1%	6	-7.3%	15
18	-7.5%	6	-9.8%	15

Note: 250 observations are used in these estimations.

Figure 3.1: *A sample of "overshooting" exchange rate due to an expansionary monetary policy shock.*
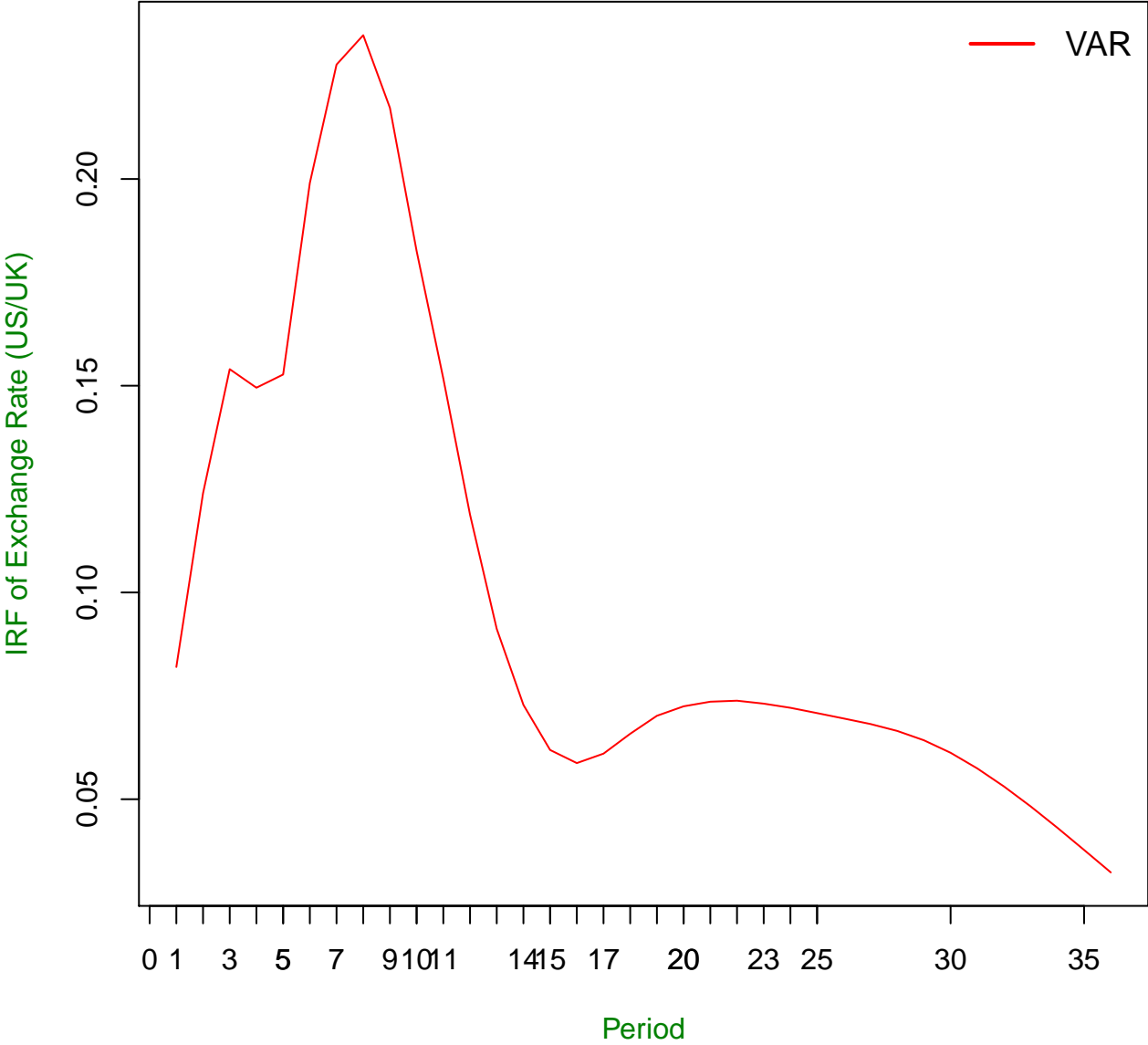
Figure 3.2: *Various Shapes of the DGP*

True DGP Peak at 3rd Period
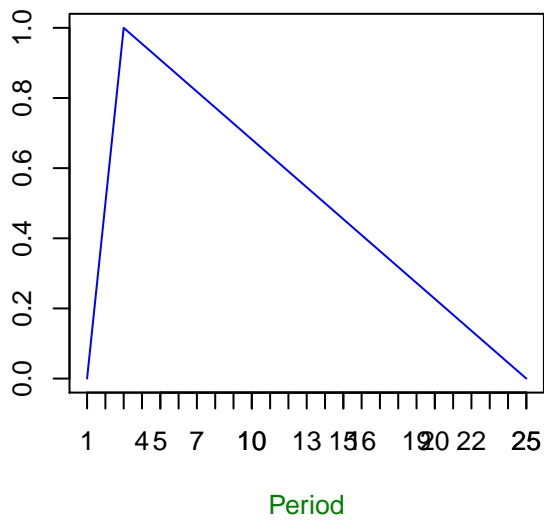
True DGP Peak at 6th Period
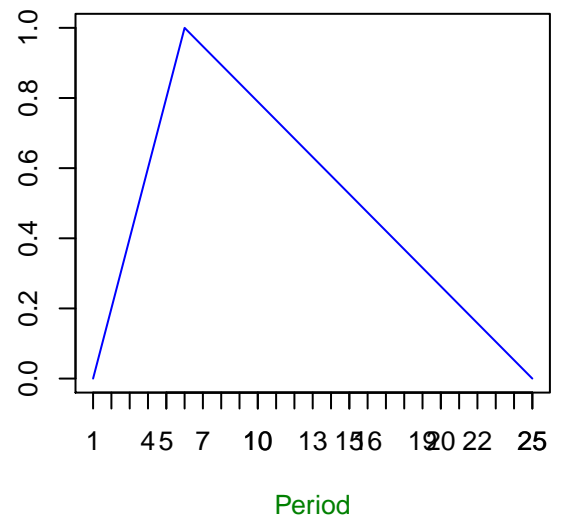
True DGP Peak at 10th Period
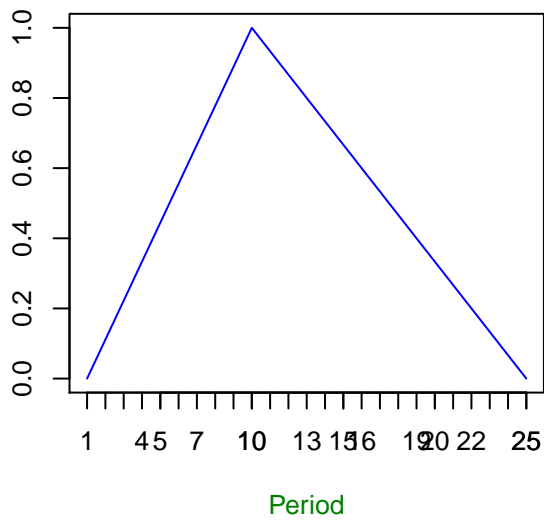
True DGP Peak at 15th Period
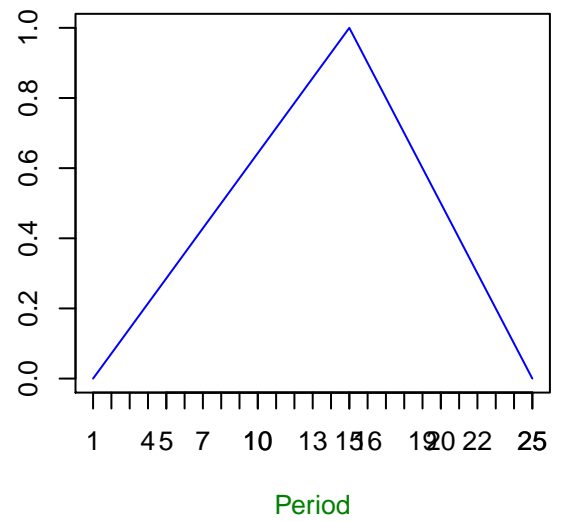
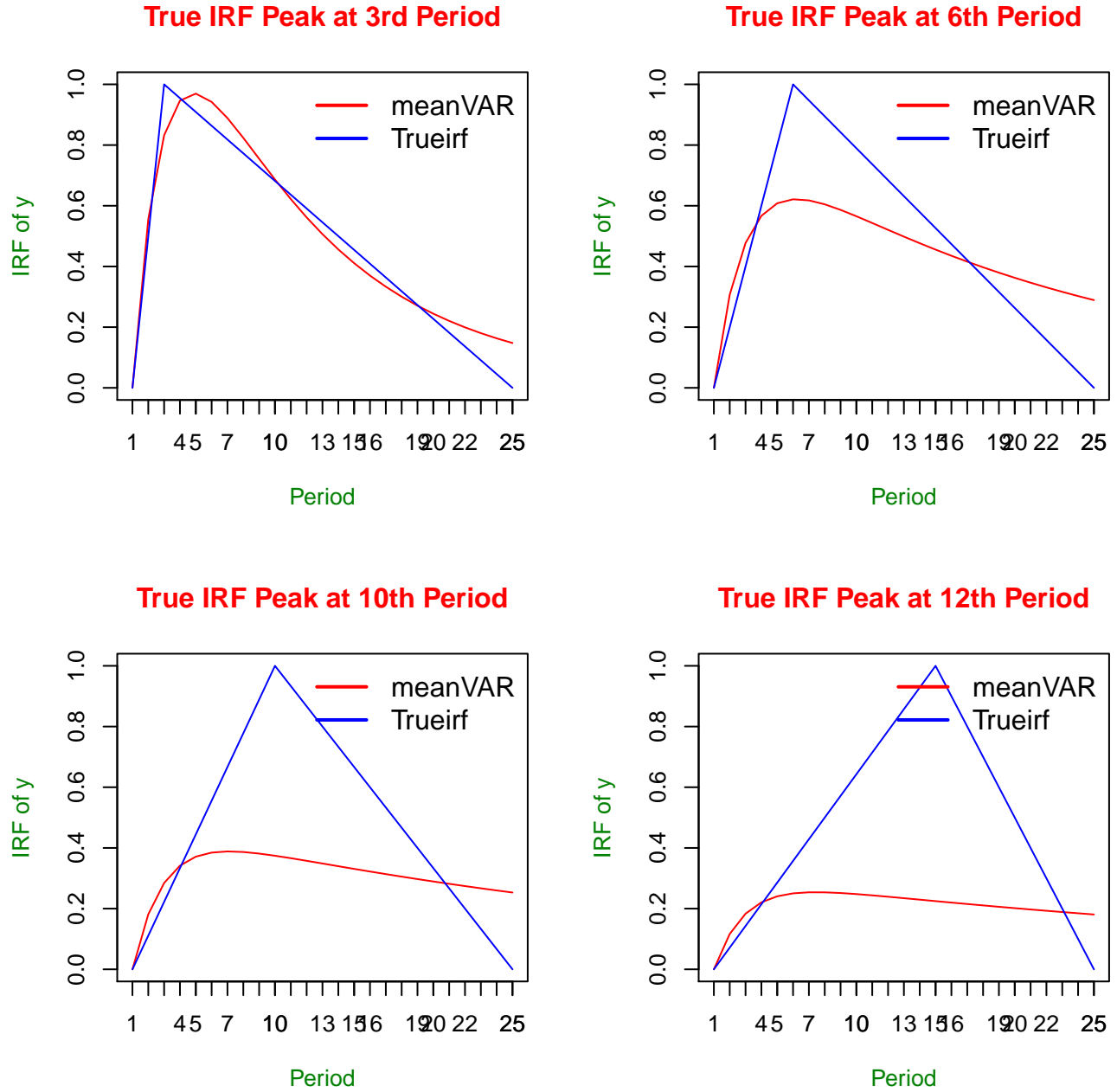**Figure 3.3**: *VAR based Impulse Response to a unit shock of x with T = 100*
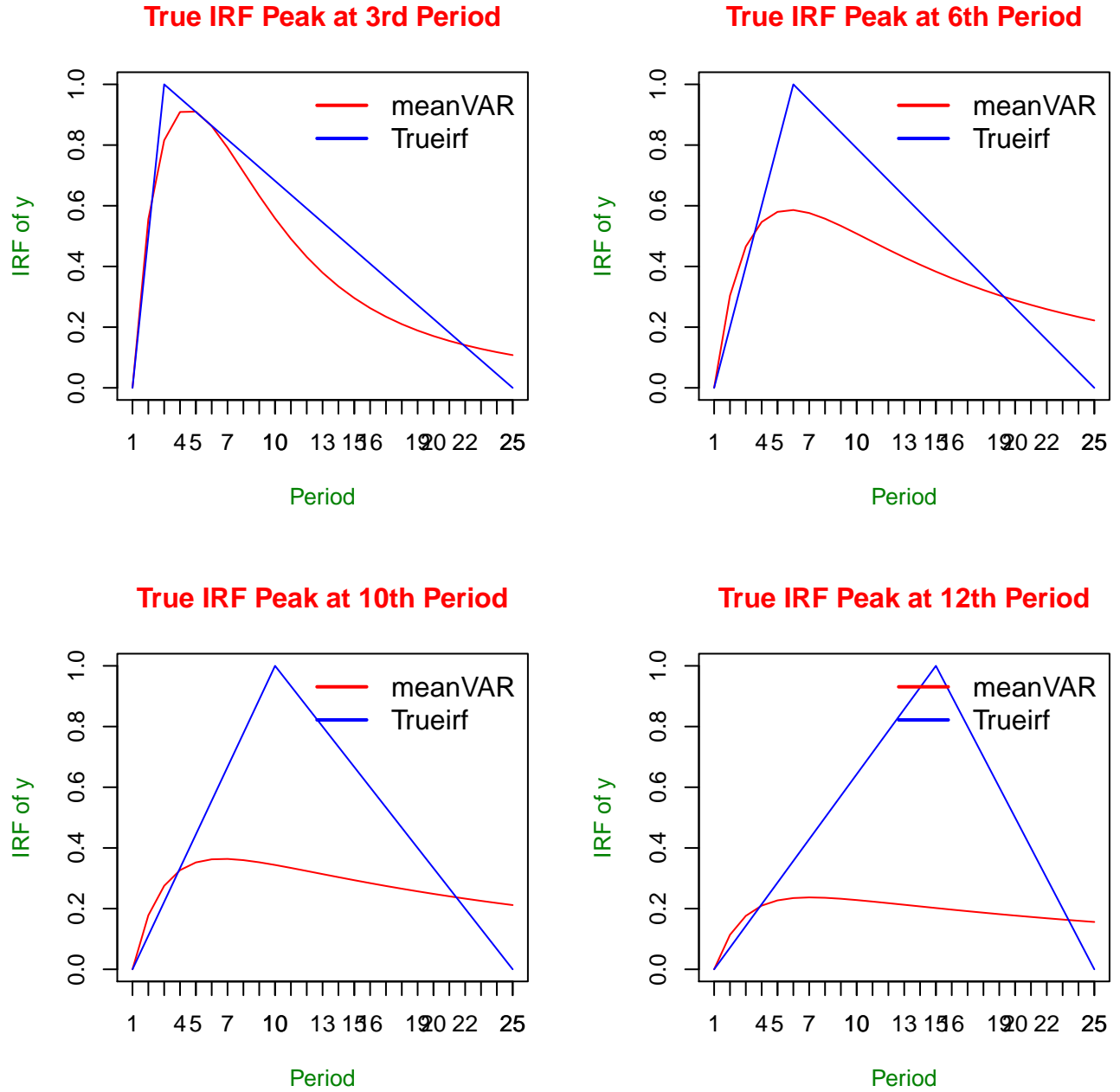
Note: The number of lags used in the VAR models is determined by the AIC.

Figure 3.4: *VAR based Impulse Response to a unit shock of x with T = 250*
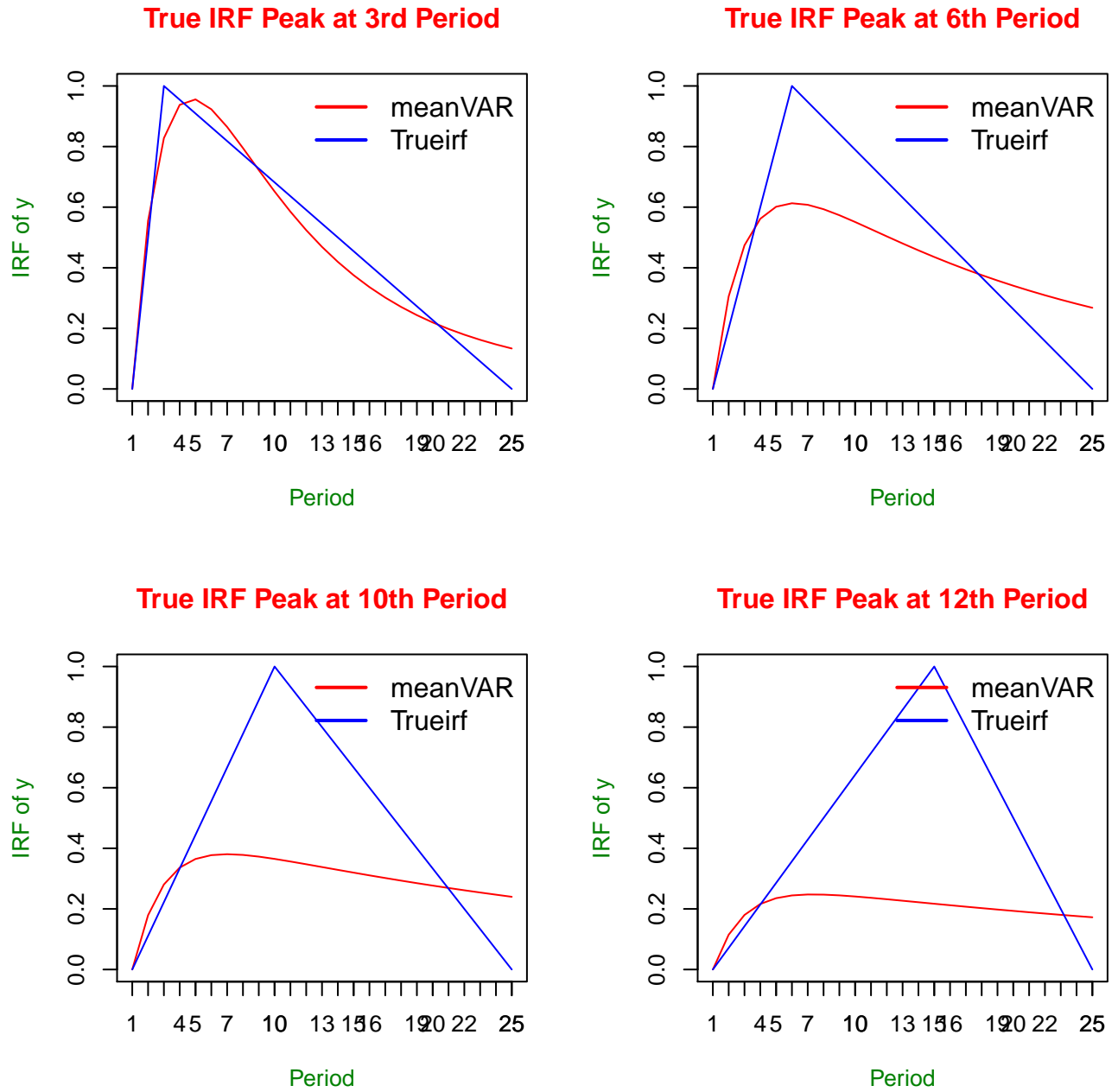
Note: The number of lags used in the VAR model is determined by the AIC.

Figure 3.5: *VAR based Impulse Response to a unit shock of x with T = 500*

Note: The number of lags used in the VAR model is determined by the AIC.

Figure 3.6: *VAR based impulse response functions with assigned number of lags and T=250*
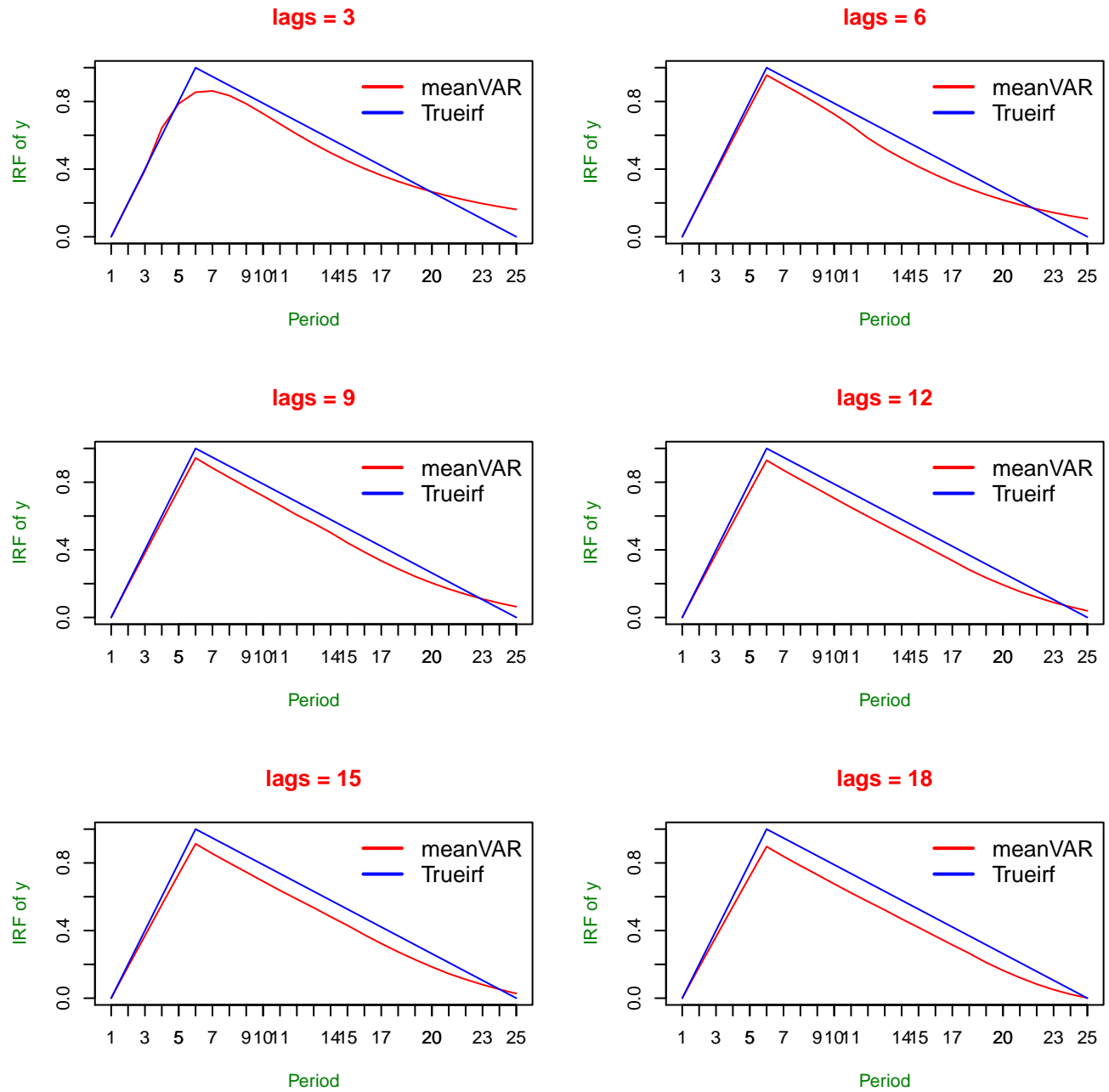
Note: The true impulse response functions peak at 6th period.

Figure 3.7: *VAR based impulse response functions with assigned number of lags and T=250*

lags = 3

lags = 6

lags = 9

lags = 12

lags = 15

lags = 18
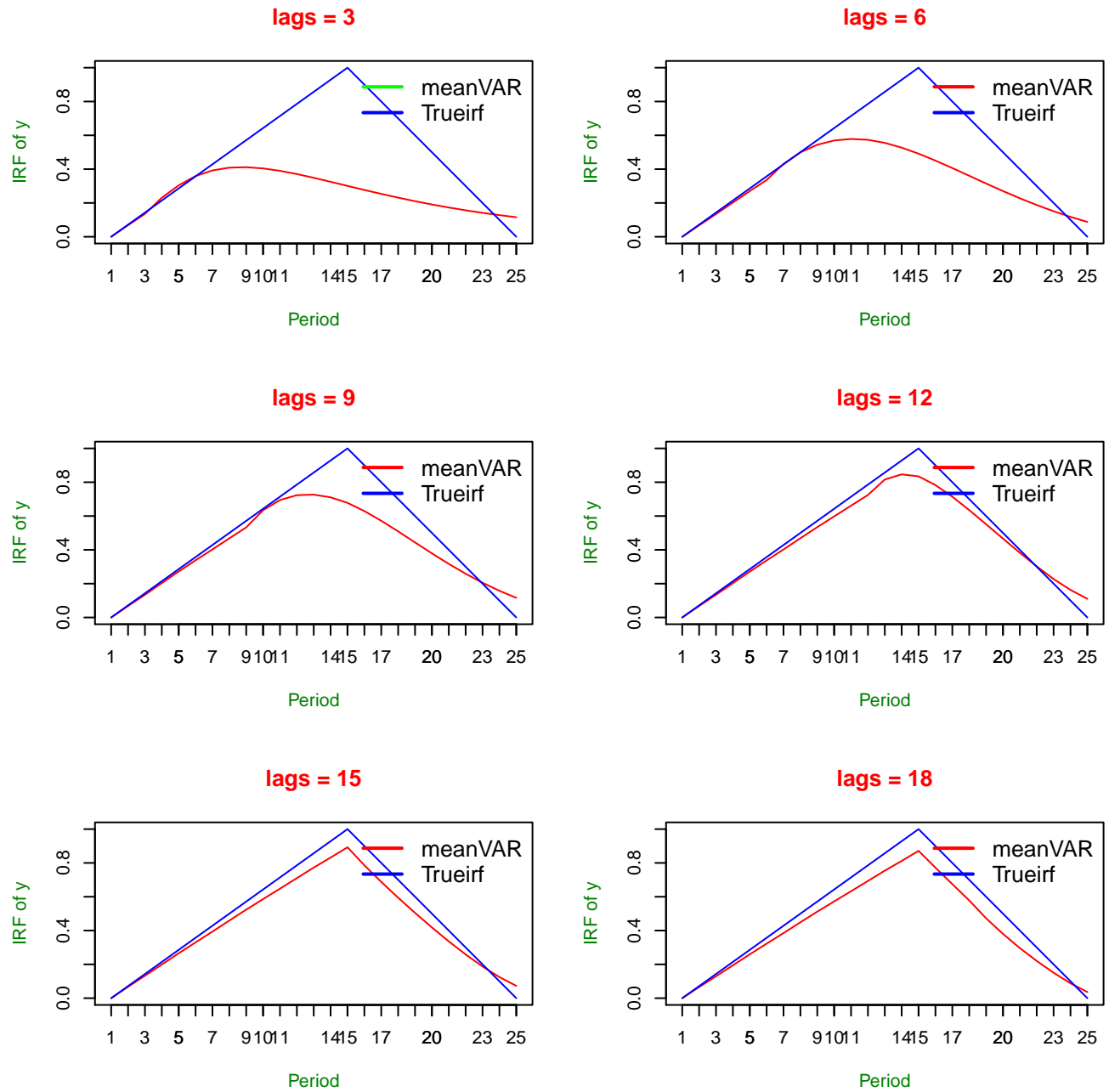
Note: The true impulse response functions peak at 15th period.

Figure 3.8: *Impulse response of y to a unit shock of x, with assigned number of lag and T = 250*
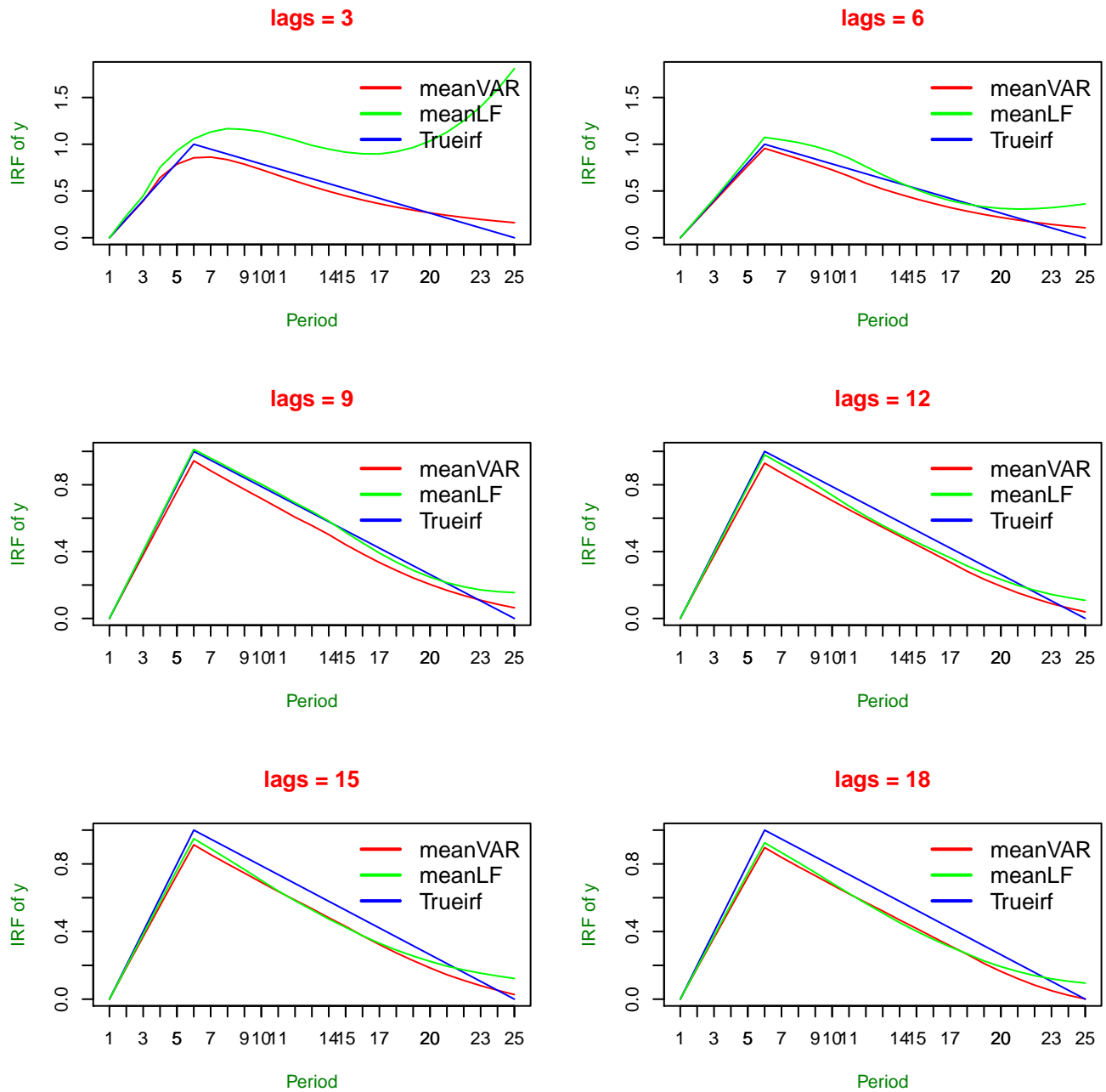
Note: The true impulse response peaks at 6th period

Figure 3.9: *Impulse response of y to a unit shock of x, with assigned number of lag and T = 250*
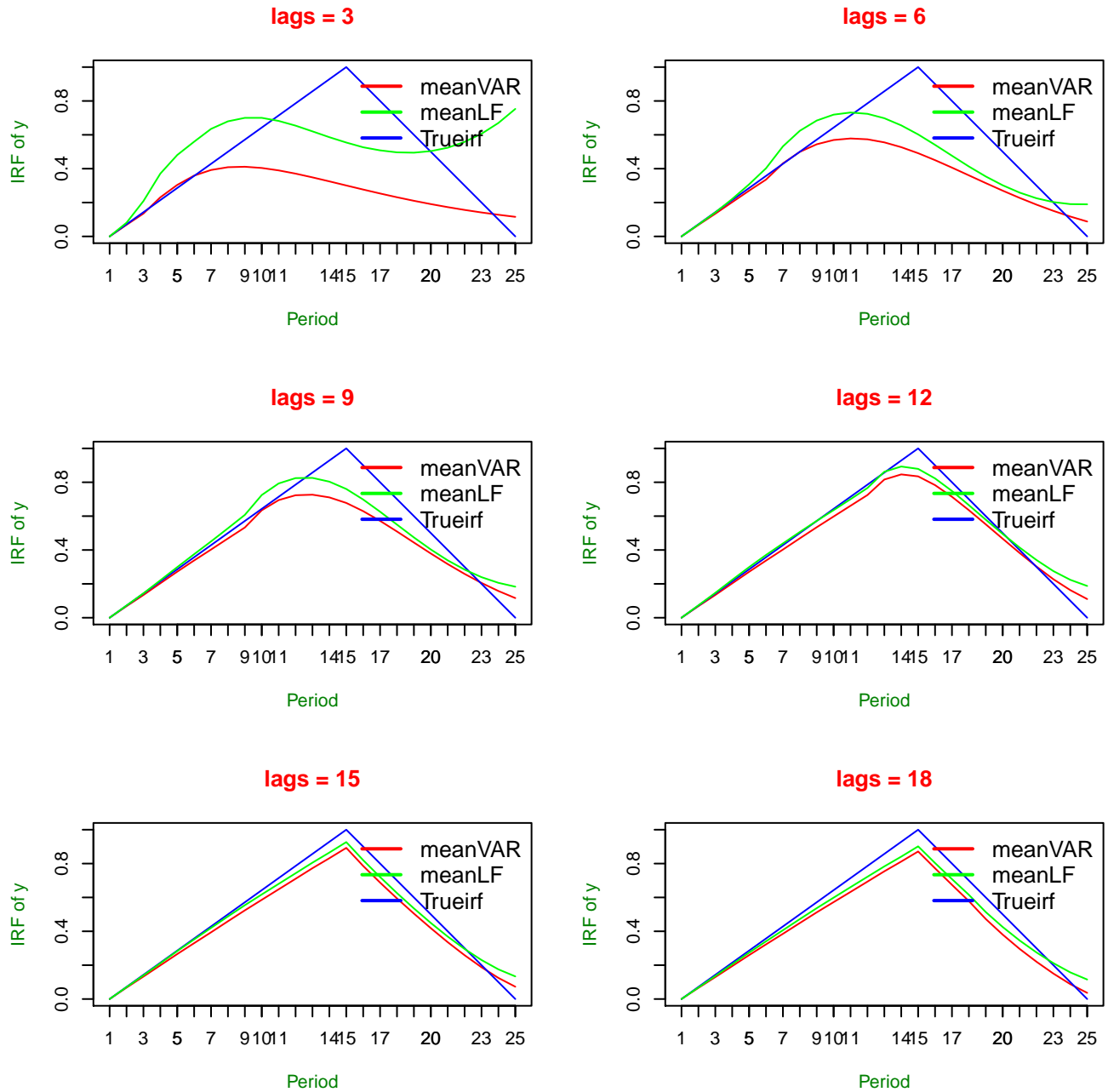
Note: The true impulse response peaks at 15th period

Figure 3.10: *Impulse response of spot exchange rate to a contractionary monetary policy shock, current spot exchange rate is above the target exchange rate*
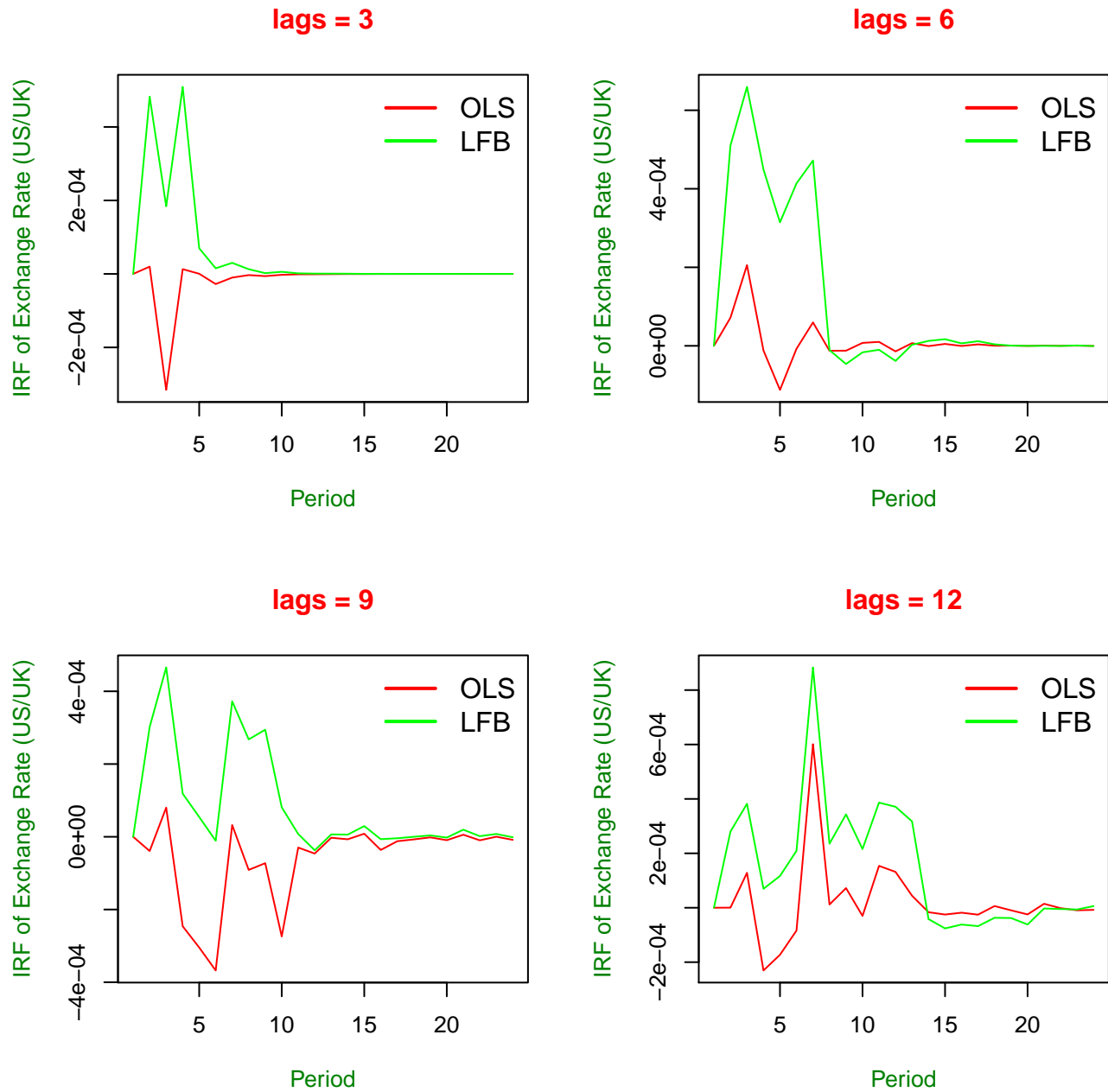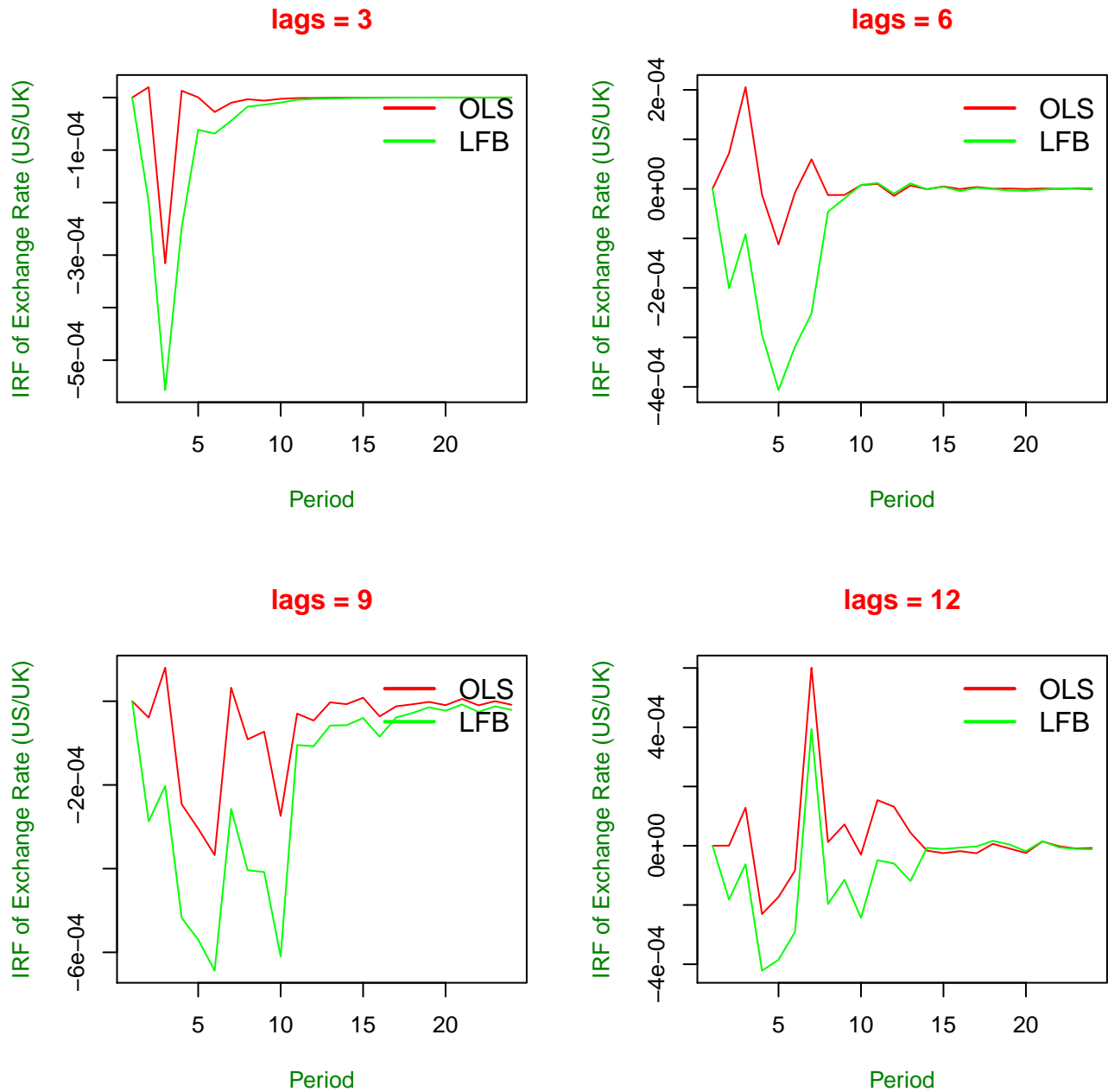

Note: 250 observations are used in each estimation and $\alpha=$ one standard deviation positive residual

Figure 3.11: *Impulse response of spot exchange rate to a contractionary monetary policy shock, with alpha=0.5 and 5, and with current spot exchange rate is above the target exchange rate*
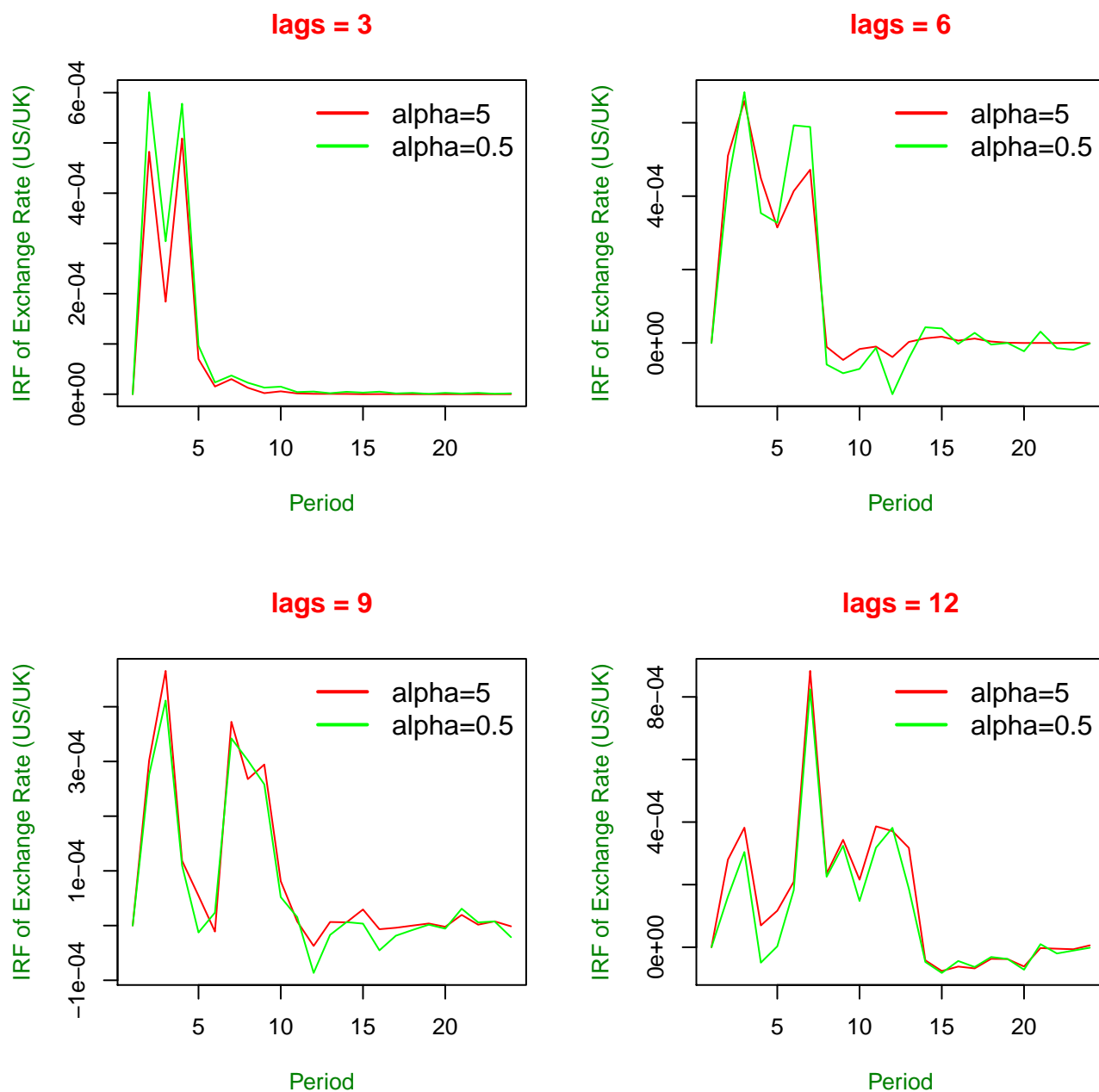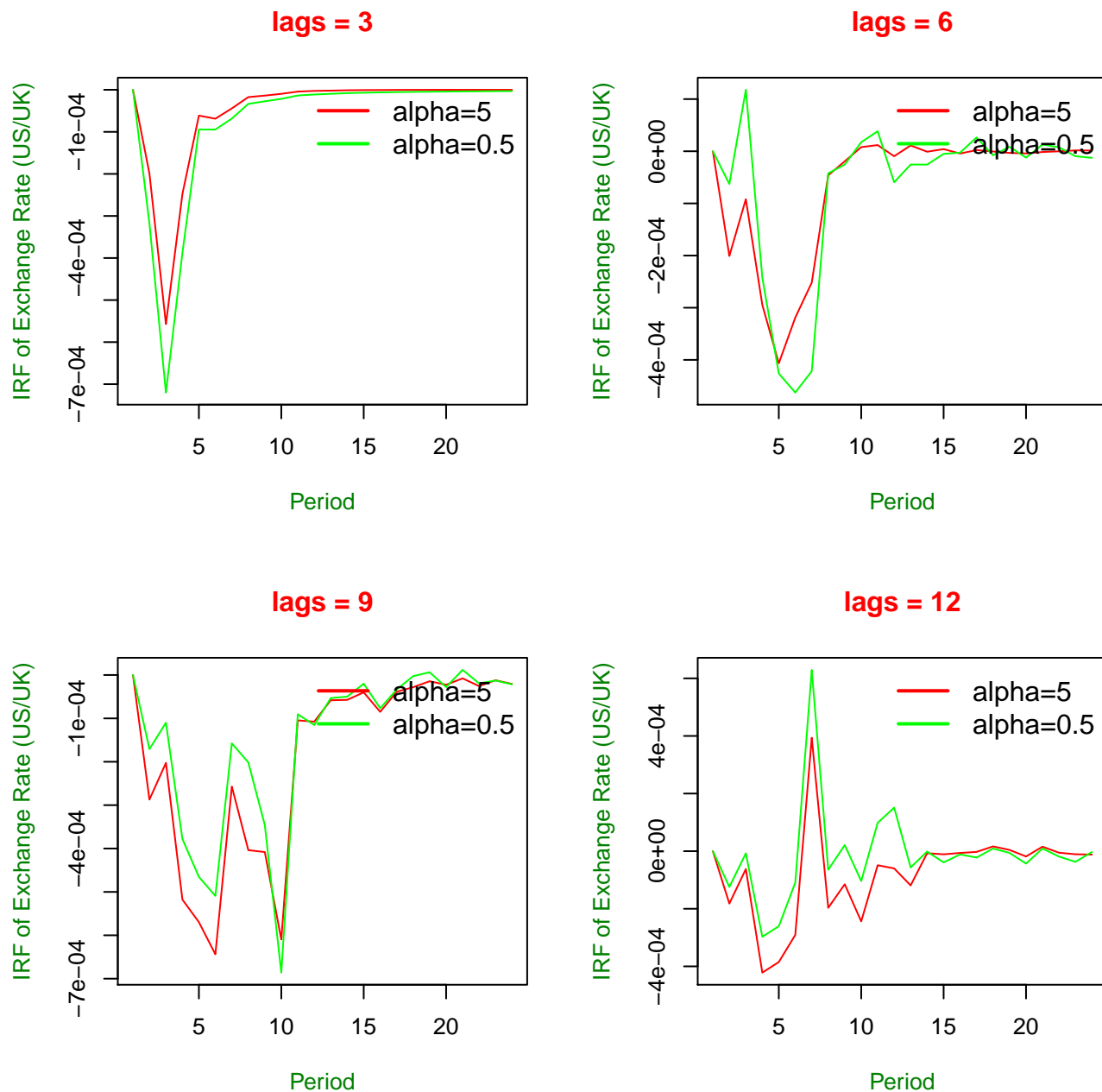

Note: 250 observations are used in each estimation and $\alpha = 0.5$ and 5

Figure 3.12: *Impulse response of spot exchange rate to a contractionary monetary policy shock, current spot exchange rate is below the target exchange rate*
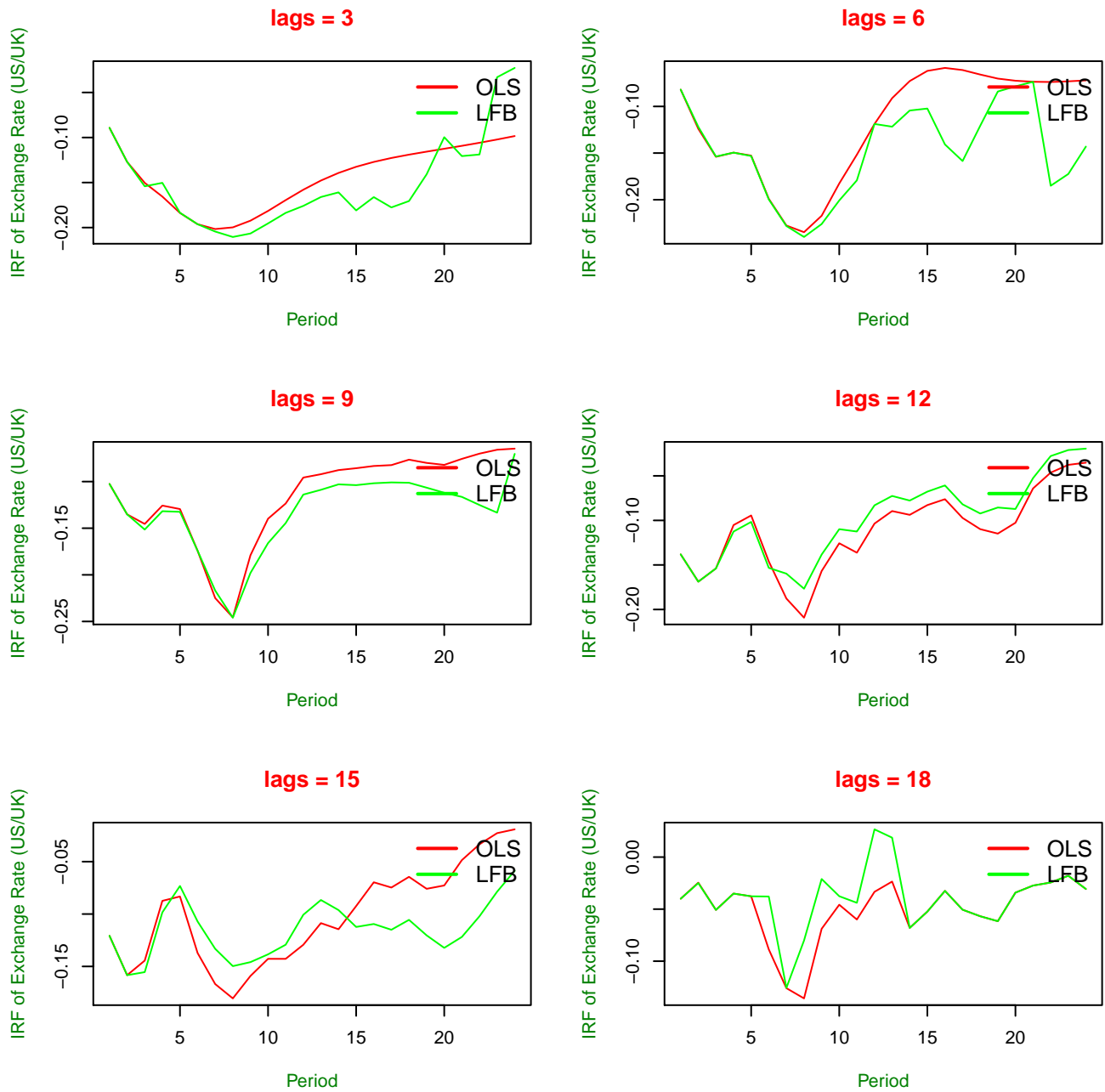

Note: 250 observations are used in each estimation

Figure 3.13: *Impulse response of spot exchange rate to a contractionary monetary policy shock, with alpha=0.5 and 5, and with current spot exchange rate is below the target exchange rate*

Note: 250 observations are used in each estimation and $\alpha = 0.5$ and 5

Figure 3.14: *Impulse response to the Exchange rate due to a contractionary monetary policy shock.*

Note: Current Exchange Rate is above the target exchange rate.

Figure 3.15: *Impulse response of spot exchange rate to a contractionary monetary policy shock, with alpha=0.5 and with current spot exchange rate is above the target exchange rate*
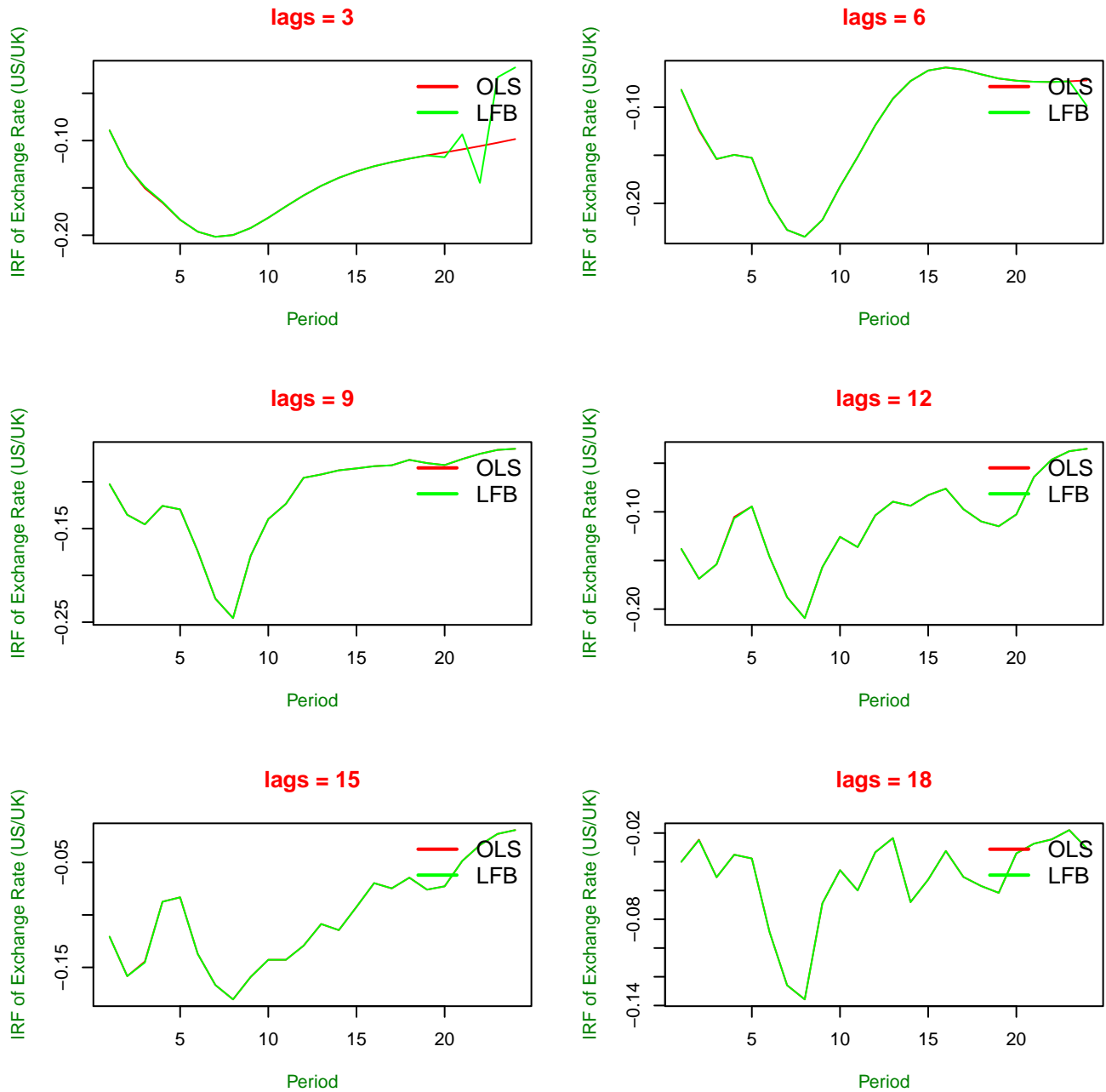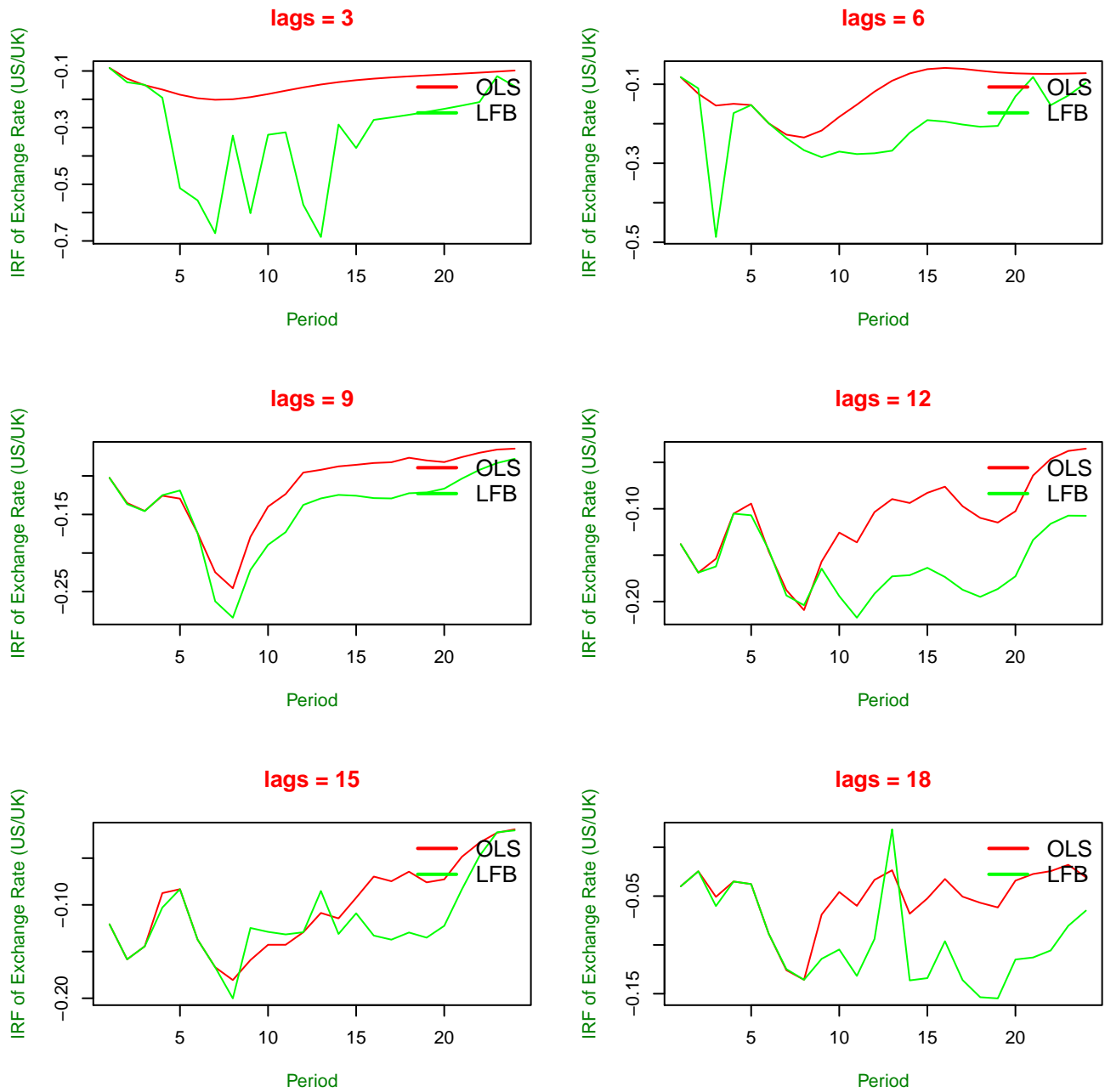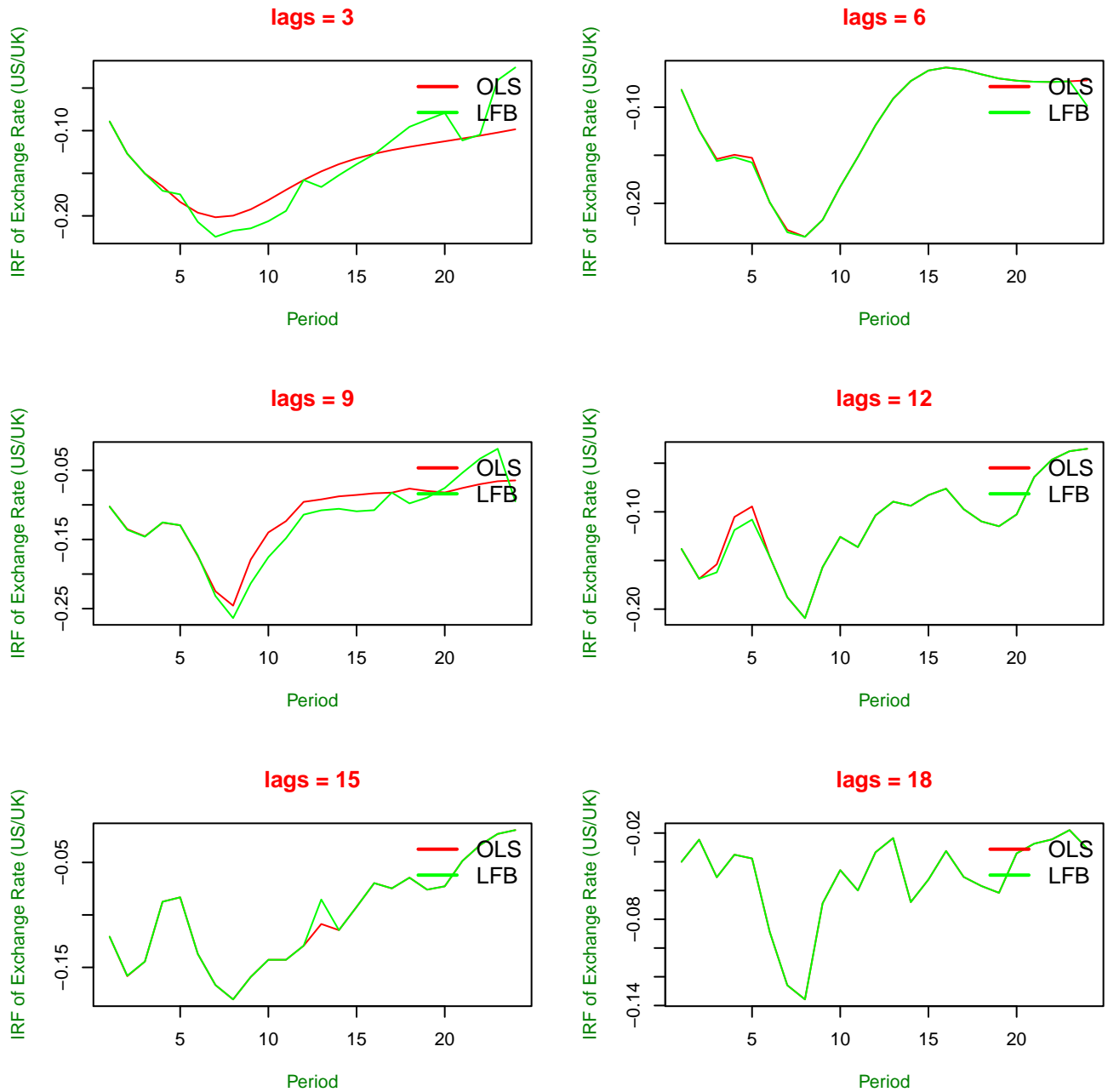
Figure 3.16: *Impulse response to the Exchange rate due to a contractionary monetary policy shock.*

Note: Current Exchange Rate is below the target exchange rate.

Figure 3.17: *Impulse response of spot exchange rate to a contractionary monetary policy shock, with alpha=0.5, and with current spot exchange rate is below the target exchange rate*

Bibliography

Us House Passes China Currency Sanctions Bill. *BBC News*, 29 Sept 2009.

Wilson defends 'pound in your pocket. *BBC News*, 19 Nov 1967.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control 19*, 716–723.

Alexander, C. and E. Lazar (2006). Normal mixture garch(1,1): applications to exchange rate modelling. *Journal of Applied Econometrics 21*, 307–336.

Anas, A. and K. Xiong (2003). Intercity trade and the industial diversification of cities. *Journal of Urban Economics 54*, 258–276.

Ball, L. (1997, March). Policy rules for open economies. Technical Report 5952, NBER Working Paper.

Bjournland, H. C. (2009). Monetary policy and exchange rate overshooting: Dornbusch was right after all. *Journal of International Economics 79*, 64–77.

Bohman, H. and D. Nilsson (2007a). Income inequality as a determinant of trade flows. *International Journal of Applied Economics 4*, 40–59.

Bohman, H. and D. Nilsson (2007b). Market overlap and the direction of exports: a new approach of accessing the linder hypothesis. Technical report, CESIS Electronic Working Paper Series.

Cavailhesa, J., C. Gaigne, T. Tabuchic, and J.-F. Thisse (2007). Trade and the structure of cities. *Journal of Urban Economics 62*, 383–404.

Clarida, R. and J. Gali (1994). Sources of real exchange rate fluctuations: how imporant are nominal shocks? In *Carnegie-Rochester Conference Series on Public Policy*, Number 41, pp. 1–56.

Clarida, R., J. Gali, and M. Gertler (1998). Monetary policy rules in practice: Some international evidence. *European Economic Review 42*, 1033–1067.

Clarida, R., J. Gali, and M. Gertler (2000). Monetary policy rules and macroeconomic stability: Evidence and some theory. *The Quarterly Journal of Economics 115*, 147–180.

Clarida, R., J. Gali, and M. Gertler (2002). A simple framework for international monetary policy analysis. *Journal of Monetary Economics 49*, 879–904.

Cushman, D. and T. Zha (1997). Identifying monetary policy in a small open economy under flexible exchange rates. *Journal of Monetary Economics 39*, 433–448.

Dai, J. and S. Sperlich (2010). Simple and effective boundary correction for kernel densities and regression with an application to the world income and engel curve estimation. *Computational Statistics and Data Analysis 54*, 2487–2497.

Dalgin, M., V. Trindade, and D. Mitra (2008, January). Inequality, nonhomothetic preferences, and trade: A gravity approach. *Southern Economic Journal 74*(3), 747–774.

Dormael, A. V. (1978). *Bretton Woods : birth of a monetary system.* MacMillan.

Dornbusch, R. (1976). Expectations and exchange rate dynamics. *Journal of Political Economy 84*, 6.

Eichenbaum, M. and C. L. Evans (1995). Some empirical evidence on the effects of shocks to monetary policy on exchange rate. *The Quarterly Journal of Economics 110*, 975–1009.

Fajgelbaum, P., G. Grossman, and E. Helpman (2009, September). Income distribution, product quality, and international trade. NBER Working Papers 15329, National Bureau of Economic Research, Inc.

Faust, J. and J. H. Rogers (2003). Monetary policy's role in exchange rate behavior. *Journal of Monetary Economics 50*, 1403–1424.

Favero, C. and M. Marcellino (2004). Large datasets, small models and monetary policy in europe. Technical report, Bocconi University.

Feenstra, R. C., R. E. Lipsey, H. Deng, A. C. Ma, and H. Mo (2005, January). World trade flows: 1962-2000. (11040).

Finger, J. and M. E. Kreinin (1979). A measure of 'export similarity' and its possible uses. *The Economic Journal 89*, 905–912.

Gavin, F. J. (2003). *Gold, Dollars, and Power - The Politics of International Monetary Relations, 1958-1971.* The University of North Carolina Press.

Grajeda, M. R. and I. Sheldon (2009). Trade openness and city interaction.

Hallak, J. C. (2006, January). Product quality and the direction of trade. *Journal of International Economics 68*(1), 238–265.

Hallak, J. C. (2010, 09). A product-quality view of the linder hypothesis. *The Review of Economics and Statistics 92*(3), 453–466.

Hamilton, J. (1994). *Time Series Analysis.* Princeton University Press.

Head, K., T. Mayer, and J. Ries (2010, May). The erosion of colonial trade linkages after independence. *Journal of International Economics 81*(1), 1–14.

Henderson, J. (1974). The size and types of cities. *American Economic Review 64*, 640–656.

Hunter, L. C. (1991). The contribution of nonhomothetic preferences to trade. *Journal of International Economics 30*, 345–358.

Hunter, W. C., G. G. Kaufman, and T. H. Krueger (Eds.) (1999). *The Asian Financial Crisis: Origins, Implications and Solutions.* Springer.

Jenkins, S. P. (1995). Did the middle class shrink during the 1980s? uk evidence from kernel density estimates. *Economic Letters 49*, 407–413.

Johnson, P. A. (2000). A nonparametric analysis of income convergence across the us states. *Economics Letters 69*, 219–223.

Kate, D. T. and R. Onsanit (2008). Thai Deputy Premier Quits as Police Battle Protesters. *Bloomberg Business October 7.*

Kearney, C. and A. J. Patton (2000). Multivariate garch modeling of exchange rate volatility transmission in the european monetary system. *Financial Review 35*, 29–48.

Kohlhagen, S. (1977). Income distribution and "representative demand" in international trade flows: An empirical test of linder's hypothesis. *Southern Economic Journal 44*, 167–172.

Krugman, P. (2010). Taking on china. *New York Times 14 March.*

Levy-Yeyati, E. and F. Sturzenegger (2005). Classifying exchange rate regimes: Deeds vs. words. *European Economic Review 49*, 1603–1635.

Lewis, D. and A. Webster (2001, JUN). Export specialisation in the caribbean and its implications for trade negotiations. *World Economy 24*(6), 809–828.

Linder, S. (1961). *An Essay on Trade and Transformation.* John Wiley, New York.

Linder Hunter, C. and J. R. Markusen (1988). *Empirical Methods for International Trade*, Chapter Per-Capita Income as a Determinant of Trade. MIT Press.

Long, N. V. and K. Y. Wong (2009). A tale of two ports: The economic geo of inter-city rivalry. *Review of International Economics 17*, 261–279.

Matsuyama, K. (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*, 1093–1120.

McDonald, J. (2011). Value of yuan likely to dominate u.s.-china trade talks. *New York Times 15 April*.

McQuillen, W. and Z. Tracer (2011, December). Chinese imports hurt u.s. solar companies, trade panel says. *Bloomberg Business December*.

Nicita, A. and M. Olarreaga (2007). Trade, production, and protection database, 1976–2004. *World Bank Economic Review 21*(1), 165–171.

Orphanides, A. (2002). Monetary-policy rules and the great inflation. *American Economic Review 92*, 115–120.

Orphanides, A. (2003). Historical monetary policy analysis and the taylor rule. *Journal of Monetary Economics 50*, 983–1022.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review 47*, 527–556.

Rauch, J. E. (1999). Networks versus markets in international trade. *Journal of International Economics 48*, 7–35.

Rodrik, D. (2008). The real exchange rate and economic growth. *Brookings Papers on Economic Activity*, 365–412.

Rudebusch, G. D. (2002). Term structure evidence on interest rate smoothing and monetary policy inertia. *Journal of Monetary Economics 49*, 1161–1187.

Sacka, B. and V. Wieland (2000). Interest-rate smoothing and optimal monetary policy: a review of recent empirical evidence. *Journal of Economics and Business 52*, 205–228.

Sala-I-Martin, X. (2006). The world distribution of income falling poverty and ... convergence, period. *The Quarterly Journal of Economics 121*, 351–397.

Schott, P. K. (2008, JAN 2008). The relative sophistication of chinese exports. *Economic Policy* (53), 6–49.

Sims, C. A. (1992). Interpreting the macroeconomic time series facts: the effects of monetary policy. *European Economic Review 36*, 975–1011.

Steelman, A. (2011). The federal reserve's "dual mandate": The evolution of an idea. Technical report, The Federal Reserve Bank of Richmond.

Svensson, L. E. (2003). Escaping from a liquidity trap and deflation: The foolproof way and others. *Journal of Economic Perspectives 17*, 145–166.

Svensson, L. E. O. (2004). The magic of the exchange rate: Optimal escape from a liquidity trap in small and large open economies. Technical Report 7/2004, HKIMR.

Taylor, J. B. (1993). Discretion versus policy rules in practice. In *Carnegie-Rochester Conference Series on Public Policy*, Volume 39, pp. 195–214.

UNU-WIDER (2008). UNU-WIDER World Income Inequality Database, Version 2.0c. Technical report.

Vilasuso, J. (2002). Forecasting exchange rate volatility. *Economics Letters 76*, 59–64.

World Bank (2011). World development indicators. Technical report, World Bank, http://data.worldbank.org/indicator Accessed June 10 .

Yellen, J. L. (2007). The asian financial crisis ten years later: assessing the past and looking to the future. In *Speech to the Asia Society of Southern California*, Volume 6 Feb.

Zheng, Y. and J. Qi (2007, JUL-AUG). Empirical analysis of the structure of sino-us agricultural trade. *China & World Economy 15*(4), 35–51.