The Foreign Exchange Market

Barry W. Ickes

Econ 434

Fall 2006

1. Introduction

The market for foreign exchange involves the purchase and sale of national currencies. A foreign exchange market exists because economies employ national currencies. If the world economy used a single currency there would be no need for foreign exchange markets. In Europe 11 economies have chosen to trade their individual currencies for a common currency. But the *euro* will still trade against other world currencies. For now, the foreign exchange market is a fact of life.

The foreign exchange market is extremely active. It is primarily an over the counter market, the exchanges trade futures and option (more below) but most transactions are OTC. It is difficult to assess the actual size of the foreign exchange market because it is traded in many markets. For the US the Fed has estimated turnover (in traditional products) in 1998 to be $351 billion per day, after adjusting for double counting. This is a 43% increase over 1995, and about *60 times* the turnover in 1977. The Bank of International Settlements did survey currency exchanges in 26 major centers and this provides some evidence. In figure 1 we present some evidence of the daily trading volume in the major cities. This shows the size and growth of the market. *Daily* trading volumes on the foreign exchange market often exceed $1 trillion,[1] which is much larger than volumes on the New York Stock Exchange (the total volume of trade on "Black Monday" in 1987 was $21 billion). The annual volume of foreign exchange trading is some 60 times larger than annual world trade ($5.2 trillion), and even 10-12 times larger than world GNP (about $25-30 trillion in 1995). You can also verify from figure 1 that the UK still accounts for the largest share of actual trades, more than 31%.

What accounts for this huge volume and its rapid growth? Although world trade has

[1] According to the BIS survey, in 1998 turnover in traditional products (spot, forwards, and fx swaps, but excluding futures, currency options, and currency swaps) was $1.49 trillion. This represented an 80% increase from 1992.

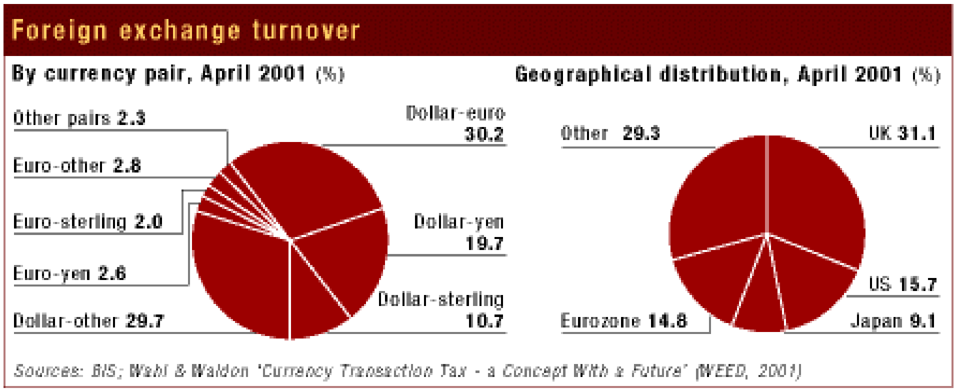

Figure 1: Foreign Exchange Turnover by Region and Currency

grown substantially – increasing 2.5 times since 1980 – this is far smaller than the growth in the foreign exchange market. International capital flows have increased more dramatically. This is related (causality is hard to infer) to increases in current account deficits is many counties, especially the US. Although the world current account must sum to zero, if the US has large deficits, other countries must have large surpluses, and this leads to an increase in international capital flows.

Moreover, there has been an expansion in international securities markets. Banks have become more multinational and more bonds are issued internationally than before. This is evident in figure 2 which shows how dramatically these have increased. This is clearly related to increasing activities of multinational companies.

Still there is a bit of a puzzle. The explanations for growth of the foreign exchange market are still too small to explain the huge volume. The reason is that the turnover in foreign exchange represents *gross* capital flows, but the explanations focus on *net* capital flows. Take the US case. Turnover is $351 billion per day, or about $87 trillion per year (assuming 250 work day). But US GDP is only about $10 trillion, and our current account deficit is "only" about 5% of that, or $500 billion. Gross transactions are thus a big multiple of net transactions. This reflects hedging behavior on the part of market participants. More below.

The vast majority of transactions in the foreign exchange market involve dollars. In 1989 the share of total turnover that involved dollars was 90%. By 1995 this had fallen to 83%.

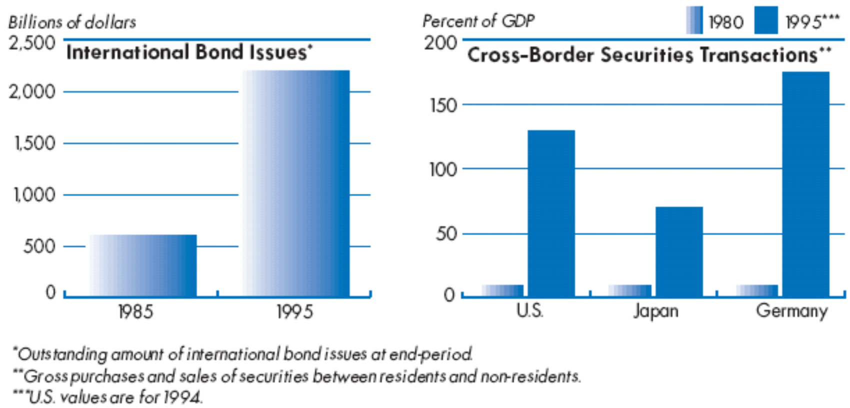

Figure 2: Growth in International Securities Markets

As of April 2001 the dollars share of total turnover has reached 90.3% (see figure 1). The next most active currency is now the Euro, which has a share of 37.6%.[2] Because of the high volume in trade in dollars (and to a lesser degree in euros and yen) many currencies do not directly trade against each other. Other pairs (those that do not include the dollar, yen, or euro) account for only 2.3% of total trade. If one wants to trade rubles for pesos it is cheaper to trade rubles for dollars and dollars for pesos, rather than engage in the direct transaction.

If you think about it, it is obvious that most trade will take place in one or a couple of currencies. Suppose that there are 150 national currencies in the world. Then, if all countries trade with each other, and if all currencies are used in these trades, then the number of foreign currency markets that would be needed is

$$\frac{n(n-1)}{2} = \frac{150(149)}{2} = 11,175. \tag{1}$$

If, on the other hand, a single currency – sometimes called a *vehicle* currency – is used on one side of all these transactions you would only need 149 markets. The savings in terms of transactions costs are enormous. Of course which currency becomes the international standard is a matter of history and chance, among other factors. But the fact that one (or two at the most) ought to predominate is most plausible.

This tremendous volume of trade is relatively recent. It has not always been the case.

[2]Because it takes two currencies to trade, these shares must add up to 200%.

Trading in currencies is much larger now than it was prior to the demise of Bretton Woods. When exchange rates were fixed there was less reason to trade.

Figure 1: Daily Volume of Trading by Location (in billions of US$)

Country	*April 1989 Turnover*	*April 1989 share*	*April 1995 Turnover*	*April 1995 share*	*Pct change 1989-1995*
United Kingdom	184	25.6%	464	29.5%	60%
United States	115	16.0	244	15.5	46
Japan	111	15.5	161	10.2	34
Singapore	55	7.7	105	6.7	43
Hong Kong	49	6.8	90	5.7	49
Switzerland	56	7.8	86	5.5	32
Germany	(NA)		76	4.8	39
France	23	3.2	58	3.7	74
Australia	29	4.0	40	2.5	37
Others	96	13.4	248	15.8	36
Total	718	100	1,572	100	46
Adjustments					
less cross-border	-184		-435		
Net-net turnover	534		1,137		45
plus estimated gaps	56		53		
= estimated global	590		1,190		45
plus futures and options	30		70		17
Grant Total	620		1,260		45

It turns out that foreign exchange trading is rather profitable. Commercial banks that engage in currency trading make rather large profits, though these are quite variable across banks and from year to year. Some evidence in table 1. The fact that many commercial banks earn large profits is rather curious. One might suspect that foreign exchange trading is a zero-sum game. Of course, traders might like to argue that these profits are due to their expertise. Can we think of an economic explanation?

One explanation for positive profits might be that the banks are providing a service for which they earn a positive return. Only the speculative activities of the banks ought to be zero sum. A back of the envelope calculation is informative. We have seen that the foreign exchange market amounts to about $1.19 trillion per day, or $300 trillion per year. Now suppose that customer trading is only 10% of total trades (actually it is larger than this) with speculative positions the remainder. Then if banks earn a fee of 2 basis points (.0002) per transaction, total profits would be $6 billion in spread income. Now a sample of the 14 largest commercial banks profits sums to about $2.1 billion. So it is quite likely that turnover income is the source of *all* profits, and that trading is actually a net loss.

Another explanation that could account for the profits from foreign exchange trading could be the activities of central banks. Central banks may engage in foreign exchange transactions that lose money, as they unsuccessfully try to defend currencies. Thus the losses of the central banks could be the source of the profits of the commercial banks. Some studies have found these losses to be very large.[3] The Fed has been rather successful since the mid 1980's,[4] but there have been some quite notable losses, most famous, perhaps, the Bank of England defending the pound in the early 1990's. The Bank of Japan similarly lost quite a lot of money trying to prevent the yen from appreciating against the dollar during 2003 and 2004.

[3] One study found that major central banks lost $16 billion on currency trading during the 1970's.

[4] The US was even more successful when it sold Carter bonds in the late 1970's. These were US debt denominated in foreign currencies. For example, the Fed sold debt denominated in DM in 1978. The interest rate on this debt was about 3% less than similar dollar-denominated debt. The reason was fear of dollar depreciation. When the debt matured, however, the dollar had appreciated, so the US earned a capital gain as well as the lower interest. You can think of this as trading on your inside information that monetary policy will be tighter than believed by the market.

Table 1: Foreign Exchange Trading Profits of Selected Commercial Banks

millions of dollars

	1985	1986	1987	1988	1989	1990	1991
Bank of America	170	141	140	135	143	207	246
Bankers Trust	107.5	57.4	512.8	153.9	296.5	425	272
Citibank	358	412	453	616	471	657	709
Morgan	172.6	229.6	251.2	186.8	190.7	309	72

The exchange rate is the price that is determined in the foreign exchange market. Of course, there are many concepts of exchange rate we can consider. These include:

- spot versus forward exchange rates versus future exchange rates
- fixed versus flexible exchange rates
- nominal versus real exchange rate

We discuss these in turn.

2. Spot and Forward Exchange Rates

The exchange rate is simply the price of foreign currency in terms of domestic currency. The typical fashion is to quote the foreign currency price of the dollar; hence, the yen has been trading at approximately Y111.925 to the dollar on September 23, 2003.[5] Similarly, the Brazilian real now trades at approximately 2.898 to the dollar, compared to the 1.2 before the exchange crisis. Of course, it is arbitrary how we quote the exchange rate; it is equivalent to say that it takes .87Euro to the dollar or that a Euro is worth 1.147 dollars. To avoid confusion we will follow the convention that the exchange rate is the price of foreign exchange in terms of domestic currency.[6] For example, the exchange rate with the Euro would be expressed as

[5] Although for some reason the value of the pound is usually quoted in terms of dollars per pound, rather than the reverse.

[6] Notice that if we refer to the exchange rate as dollars per DM, it means that an appreciation of the exchange rate is equivalent to a depreciation of the dollar. When e goes up it takes more dollars to buy a DM, which means the dollar is worth less.

the number of dollars per Euro, because for the US the Euro is foreign exchange. Often, we will simply treat the rest of the world as one country, and hence we will simply refer to foreign exchange rather than specify the specific country. In that case the exchange rate, e, is just the domestic currency price of foreign exchange:

$$e = \frac{\text{domestic currency}}{\text{foreign exchange}}$$

The *spot* (or nominal) exchange rate refers to the current price of foreign exchange. It is a contract for immediate delivery, though that might actually take a day or two. A *forward* contract refers to a transaction for delivery of foreign exchange at some specified date in the future. Suppose, for example, that a US company, say Ford Motors, expects to receive $Eu100,000$ 60 days from now. The value of these receipts will vary with the actual value of the spot exchange rate in the future. The firm may wish, however, to hedge. It may wish to reduce the risk that the dollar will appreciate during these 60 days. Consequently, it signs a contract to deliver $Eu100,000$ in 60 days at the current exchange rate. The company has locked in the current rate and hedged the exchange rate risk. Similarly, if Microsoft commits to invest $GBP1m$ in 6 months time, it may wish to fix the dollar amount of this investment now. Hence, it could purchase a contract today to deliver pounds six months from now. Suppose that the forward price of pounds in this transaction is \$1.637.[7] Then Microsoft pays \$1,637,000 today to obtain $GBP1m$ in 6 months time. If the pound appreciates this is a profitable transaction.[8] Even if it does not, Microsoft has reduced its risk.

So far we have only analyzed Microsoft's interest. But the risk that Microsoft has hedged the bank has absorbed. But the bank actually is acting like a dealer. Simultaneously, it will be looking for other agents who need to hedge against the dollar depreciating. For example, suppose that Coca Cola expects to earn $GBP1m$ in six months from exports to the UK. It will then have to convert the pounds into dollars. To avoid the currency risk Coca Cola would

[7] The spot price on September 21, 2003 was 1.6479.

[8] For example, if the pound happened to trade at \$1.73 in six months, then Microsofts hedge would have saved them \$93,000 minus transactions fees.

like to sell the pounds forward: accept dollars today for the commitment to deliver pounds in six months. Notice that Coke is hedging against the risk that the pound will depreciate, while Microsoft is hedging the risk that the pound will appreciate. Trade is thus mutually beneficial; the bank is merely the intermediary.[9]

In this example, the intentions of Coke and Microsoft exactly balance. More typically, a bank will have many clients whose interests differ. The intention of the bank is to balance the two sides of the market and profit from the fees.

Notice that the forward price of a currency need not be equal to the spot price. If the market expects that the franc will depreciate over the next 6 months, the forward price will be lower than the spot price. The *forward premium* is a measure of the market's expectation, and it can be expressed as:

$$f_m = \frac{F_m - e}{e} \tag{2}$$

where m is the number of days from today and F_m is the forward exchange rate. Clearly, if $f_m > 0$ it means that more dollars will be required to purchase foreign exchange m days from now than today.

Forward contracts are usually offered by commercial banks, and this helps to explain the difference with *futures* contracts. Banks offer their important customers forward contracts as part of their business relationship. It enables firms to engage in international trade with limited exposure to foreign currency risk. Now in the late 1960's many observers expected the British government to devalue the pound. Milton Friedman wanted to bet on this, and he tried to purchase forward contracts to sell the pound short; that is, he would receive dollars today for the obligation to deliver pounds in several months time. If the pound were devalued he would be able to purchase the pounds for fewer dollars in the future, and hence would profit

[9] Notice that there is no need for Coke and Microsoft to have different expectations for this trade to be profitable if the firms are risk averse. Risk averse agents are willing to pay to obtain certain future outcomes. If the firms were risk neutral, on the other hand, then for trade to occur they would need to have different expectations about the future value of the pound. Thus, if Coke expected depreciation and Microsoft expected appreciation, they would both prefer to hedge, even if they were risk neutral.

This would not be the case for risk lovers. In that case some differences in expectations are needed for trade. To see this, consider betting on sports events. If everyone believed the outcome of the game would be the same, no one would bet.

on this transaction. To his surprise, however, Friedman found no banks that were willing to make this contract. The reason is that banks only sold forward contracts to their commercial clients. Friedman then was instrumental in creating the futures market.

A futures market is a clearinghouse where goods for future delivery are sold. In such markets standardized commodities are sold. In a forward exchange transaction the names of the parties and amounts are negotiated. In futures contracts, the commodity is standardized, and anyone can buy or sell it. The clearinghouse acts as the third-party guarantor. This means that agents trading in the futures market do not need to know the identity of the agents they trade with; the exchange stands behind the trades. In the forward market, on the other hand, the contract is between the bank and the firm, so the bank requires information about the creditworthiness of the customer to protect themselves against default.

If banks protect against default by knowing their customers, how does the exchange stand to guarantee trades? Futures markets rely on margin requirements. These are best thought of as deposits, or bonds, placed by traders with the exchange. Buyers and sellers post bonds with the exchange as guarantees against default. Futures contracts are marked to market. This means that the daily gains and losses on outstanding contracts are recorded and either added or deducted from an investor's margin account. If a margin account falls below some critical level the exchange requires replenishment.[10]

Margin requirements represent a key difference between forward and futures markets, and this is why most firms use the forward market (it is cheaper). The futures market is dominated by speculators. Forward markets are much larger, but futures markets are more liquid. Of course, there is arbitrage between the markets so the prices cannot deviate very far from each other.

In addition to futures contracts, one can also purchase options. The key difference between a futures contract and an option contract is that the former is an *obligation* to deliver something in the future, while the latter confers the opportunity. If Microsoft is not certain it will make the investment in 6 months, it may prefer to purchase an option rather than a fu-

[10] Unless you are Hillary Clinton.

tures contract. Microsoft would be paying some dollars today for the right to purchase francs within some future period at the pre-arranged exchange rate. A call option, for example, would provide Microsoft the right to purchase francs at the pre-arranged price during the life of the option. If the value of the franc increased holding the option would benefit Microsoft. If the value of the franc decreased, Microsoft would choose not to exercise the option. It would lose what it paid for the option, but it would not have locked itself into an unprofitable transaction. An option contract thus includes a measure of insurance. It allows the investor to avoid potential losses that would accompany unfavorable movements in currencies.

2.1. Covered Interest Parity

The existence of forward markets for foreign exchange benefits not only firms that expect to have foreign currency transactions in the future, but also investors who wish to invest in foreign currencies. Suppose that the domestic interest rate (on say, 3 month T-bills to provide specificity) is given by i, and that the foreign interest rate is i^*, and that $i^* > i$. I might wish to invest in foreign assets rather than domestic assets. But if I do so, I would face the risk that 3 months from now the dollar may appreciate, so that when I convert my foreign currency into dollars I would take a capital loss. The forward market allow an alternative, called a covered transaction, which eliminates currency risk.[11]

Letting e_t be the spot exchange rate (dollars per euro) and F_t the current forward price for euro three months hence, I could hedge my risk by purchasing a forward contract. Specifically, assume that I choose to invest a dollar. I can convert this into $\frac{1}{e_t}$ euro. I then invest this, and at the end of three months I have $\frac{1}{e_t}(1+i^*)$ euro. Because I purchased the forward contract, this yields me $\frac{F_t}{e_t}(1+i^*)$ dollars. This transaction is called *covered* because I have already closed the transaction. The covered transaction indicates the dollar return to investing in

[11] Notice that a *covered* transaction implies no currency risk. So arbitrage will equalize returns even if agents are risk averse.

This is certainly true for US and German transactions. But in the summer of 1998 ruble forward transactions were anything but certain. And for good reason; most sellers of forward contracts (in dollars) were unable to honor their contracts. Because of high ruble yields on Russian T-bills (called *GKO's*) foreign investors flocked to Russia and purchased forward contracts to convert profits back into dollars. But in the wake of the currency crisis, the banks could not buy the dollars needed to honor the contracts.

foreign assets. Of course, I could always invest in domestic assets and earn $(1+i)$. Hence, arbitrage should insure that

$$1+i=\frac{F_t}{e_t}(1+i^*) \tag{3}$$

which is called the *covered interest parity* condition (CIPC).[12] A host of studies have shown the high degree to which this condition is satisfied by market rates.

The *CIPC* can be used to reveal an interesting relationship between interest rates and the exchange rates, by using the expression for the forward premium (2). From the definition of f_m it follows that

$$\frac{1+i}{1+i^*}=\frac{F_t-e_t}{e_t}+1=f_m+1. \tag{4}$$

In other words, when the forward premium is positive domestic interest rates are higher than foreign interest rates, and vice versa. The forward premium reflects the capital gain on my covered transaction, and by arbitrage, this must equal the difference in interest rates in the two countries.

It is easier to interpret (4) if we take logs of (3) to obtain:

$$\log(1+i)=\log F_t-\log e_t+\log(1+i*)$$

and then use the fact that for small x, $\log(1+x)\approx x$,[13] we obtain:

$$i=\log F_t-\log e_t+i^*$$

[12] In practice this is an approximation, because of transaction costs.

[13] This follows from taking a first-order Taylor approximation to x, which yields: $x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+...\approx x$.

or[14]

$$i - i* = \log F_t - \log e_t = f_t \tag{5}$$

which says that the interest differential is essentially equal to the forward premium.[15] The market's guess about the path of exchange rates thus affects current interest rates. Expectations matter.

2.2. Uncovered Interest Parity

We can obtain a further relationship between interest rates if we consider the *uncovered* version of the transaction. An investor always has the option of investing in an uncovered fashion by neglecting to purchase the forward contract. What is the expected return to this transaction? Let $\widehat{e}_{t+1}$ be the exchange rate expected to hold when the investment expires. Then the expected return to by uncovered investment equals $\frac{\widehat{e}_{t+1}}{e_t}(1 + i*)$, so that arbitrage requires:[16]

$$1 + i = \frac{\widehat{e}_{t+1}}{e_t}(1 + i*) \tag{6}$$

which is called the *uncovered interest parity* condition (UIPC). We can again take logs of both sides of (6) which yields:

$$i - i* = \log \widehat{e}_{t+1} - \log e_t = \frac{\widehat{e}_{t+1} - e_t}{e_t} \equiv \delta_t \tag{7}$$

where we have defined δ_t as the expected depreciation of the domestic currency.

It is useful to compare expression (7) with the CIPC condition, which looks similar. Note the difference between covered and uncovered interest parity. In the former there is no currency

[14] We could alternatively derive the (5) by taking logs of (4):

$$\log(1 + i) - \log(1 + i^*) = \log(1 + f_m)$$

and again using the fact that for small x, $\log(1 + x) \approx x$ we obtain:

$$i - i^* = f_m.$$

[15] Recall that $f_m = \frac{F_m - e}{e}$, so that $1+f_m = \frac{F_m}{e}$. Hence, if we take logs of both sides we get the approximation: $f_m = \log F_m - \log e$.

[16] If investors are risk neutral; that is, they care only about expected yields.

risk, hence arbitrage occurs whether or not agents are risk averse. Uncovered interest parity, on the other hand, results only if agents are risk neutral.[17] If agents are risk averse, then they will demand a risk premium to hold the risky return. In this case the only differential risk is the currency risk associated with holding foreign assets. Hence, arbitrage would result not in (6), but rather in:

$$1 + i = \frac{\widehat{e}_{t+1}}{e_t}(1 + i*) + \rho_t \tag{8}$$

where ρ_t is the risk premium.

Finally, notice that if expected future exchange rates are equal to forward rates (i.e., $F_t = \widehat{e}_{t+1}$) it follows that the UIPC must hold. It is an important item of research to test for this. Although we often assume it, in practice UIPC tends not to be supported by the data. If risk premia are important, then the forward rate is not equal to the expected exchange rate. Rather we have $f_t = \delta_t + \rho_t$. The forward premium differs from the expected rate of currency depreciation by the risk premium. In this case we cannot recover the market's expectation about the exchange rate directly from interest rate differentials.[18] We will return to this later.

Testing for Uncovered Interest Parity How can one test whether the UIPC holds? Notice that if both conditions held it follows that $F_t = \widehat{e}_{t+1}$. We have data on the former but not on the expected spot rate. To form a test, we therefore need an hypothesis about exchange rates. Hence, to test for UIPC we then impose the *rational expectations hypothesis.* This says that the expected value of a variable is given by the conditional expectation of that variable using the appropriate economic model. For our purposes the key implication is that expectations will be *unbiased*. This implies that $\widehat{e}_{t+1}$ will be an unbiased predictor of e_{t+1}.

Now we can test for UIPC. If UIPC and rational expectations hold, then F_t should be an unbiased predictor of e_{t+1}. I can always collect a time series of spot and future exchange

[17] At the margin.

[18] This also explains why investors demand a large premium to hold Brazilian reals (in February 1999) even when the real has severely depreciated. Presently, interest differentials are close to 36%, while the real has probably reached a trough relative to the dollar. But investors are worried that Brazil may default on its debt, so a large risk premium is required to induce agents to hold Brazilian assets.

rates. A regression of the form:

$$e_{t+1} = \alpha + \beta F_t + \gamma X_t + \varepsilon_t \tag{9}$$

should, according to our hypotheses result in estimates of $\widehat{\alpha} = 0, \widehat{\gamma} = 0$ and $\widehat{\beta} = 1$, where X_t are any other variables we may include in the regression. Tests usually reject this hypothesis, typically finding $\widehat{\beta} < 1$.

For technical reasons – the fact that e and F have a common trend – it is preferable to test this in a different manner. Note that UIPC implies that $\widehat{e}_{t+1} - e_t = i - i*$. From CIPC we can replace the interest differential with $F_t - e_t$. Hence, we have $\widehat{e}_{t+1} - e_t = F_t - e_t$. Of course we still cannot observe $\widehat{e}_{t+1}$, but rational expectations implies that $e_{t+1} - \widehat{e}_{t+1} \equiv \varepsilon_{t+1}$ (the forecast error) will be uncorrelated with any information known at time t. Hence, we can write

$$e_{t+1} - \varepsilon_{t+1} - e_t = F_t - e_t$$

which can be tested by estimating

$$e_{t+1} - e_t = \alpha + \beta(F_t - e_t) + \gamma X_t + \varepsilon_{t+1} \tag{10}$$

where the null is that $\widehat{\alpha} = \widehat{\gamma} = 0$ and $\widehat{\beta} = 1$.

Notice that this is then a joint test: we are testing the assumption that the forward rate is an unbiased estimator of the spot rate – i.e., that agents have rational expectations – and that agents are risk neutral. A rejection of the hypothesis could thus arise from one of two reasons.

1. expectations are not rational
2. risk neutrality

One way to view the evidence is to look at a plot of F_t and e_{t+1}. According to the hypothesis the difference between these two variable should be random error. But the plot shows in fact that the differences are systematic. Indeed, the spot rate seems to lead the forward rate. This is evident in figure 3 which plots the Yen forward and spot rate versus the dollar.

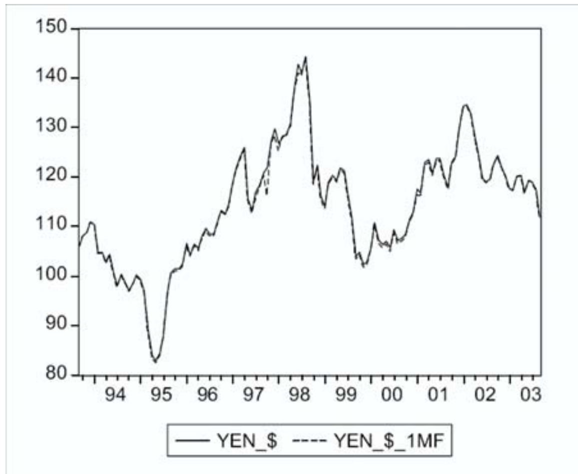

Figure 3: Yen Forward and Spot Rate

Another way to look at this is to look at the relationship between the actual change in the exchange rate $(e_{t+1} - e_t)$ and the forward discount $(F_t - e_t)$. We might expect that the latter would be less noisy than the former, but that there should be no systematic relationship. But when we plot this we see the noise but also that the deviations do not appear random. This is evident in figure 4 which plots this relationship.

Notice that if UIPC does not hold one perhaps can make money. How? One can employ the carry trade: borrow in the currency that has lower interest rates. Let y_t be the amount of dollars borrowed. Then this strategy implies

$$y_t = \begin{cases} > 0 & \text{if } i_t < i_t^* \\ < 0 & \text{if } i_t > i_t^* \end{cases} \tag{11}$$

and the payoff to this strategy is

$$y_t \left(e_t(1 + i_t^*) \frac{1}{e_{t+1}} - (1 + i) \right). \tag{12}$$

Suppose, for example, that $i > i^*$. Then if we borrow abroad and invest at home we should earn money since the actual change in the exchange rate is less than what is required by the arbitrage condition. My profit in this case should be:

$$p^+ = (1 + i) - (1 + i^*) \frac{e_{t+1}}{e_t}; i - i^* \geq 0$$

and when $i < i^*$ I follow the opposite route. My profits in this case are:

$$p^- = -\left((1+i) - (1+i^*)\frac{e_{t+1}}{e_t}\right); i - i^* < 0$$

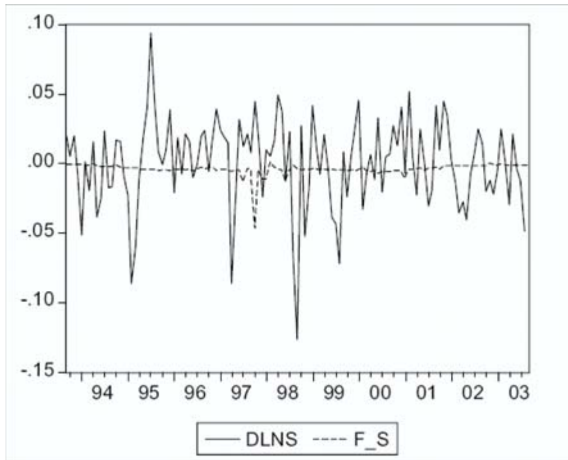

Figure 4: Realized exchange rates and the forward premium

Another version of the carry trade would be to sell domestic currency forward (going short on domestic currency) whenever there is a forward premium ($F_t > e_t$) and buying the domestic currency forward (going long on domestic currency) whenever it is at a forward discount ($F_t < e_t$). That is,

$$x_t = \begin{cases} > 0 & \text{if } F_t > e_t \\ < 0 & \text{if } F_t < e_t \end{cases} \tag{13}$$

where x_t is the number of dollars sold forward. The payoff to this strategy is

$$x_t\left(\frac{F_t}{e_{t+1}} - 1\right). \tag{14}$$

If UIPC holds, then this strategy (13) yields positive profits whenever the conventional carry trade (11) yields positive profits, and vice versa. This might be more profitable, however, as there are fewer transactions costs.

Suppose you followed this strategy (this is sometimes called the carry trade). Would you make money? Suppose you did this from September 1993 to August 2003. Bet each month for

Date	f-s	#yen/$
1998:8	-0.001005	144.3200
1998:9	-0.004189	135.1500
1998:10	-0.004547	119.0400

Figure 5:

the ten years, so you have 120 observations. Turns out you would make money. The average value of $p = .0041$. We can plot the distributions of realized profits in figure 6. You can see that they are volatile, but you do make money. But the profits are risky. The standard deviation of $p = .033$, which implies that the Sharpe ratio ($\frac{mean}{std}$) is 0.12.[19] This appears risky; the Sharpe ratio for the S&P 500 is about .06.

LTCM and Tiger Management Fund made this type of bet and lost in 1998:8 and 1998:9.[20] Were they unlucky? Here are the numbers. We use covered forward interest parity for the interest differential since that does hold.

In August 1998 the monthly interest rate in the US was 0.1% higher than in Japan. So invest in the US. Bad move, the dollar depreciated by 7% (yen appreciated 7%) and LTCM lost 6.4% (a 2 std. deviation event) on the bet.[21] And September was even worse. The interest differential was 0.4% in favor of the US, but the dollar depreciated by 13% (yen appreciated 13%,) and LTCM lost 12%, (a 3.5 std. deviation outlier, and the minimum profit in the sample). The estimate is that Tiger lost $2 billion on one day in October on such trades.

Suppose that we find that $\widehat{\beta} < 1$.[22] What does this mean? It is important to recall that

[19] The Sharpe ratio is the excess return on an asset divided by the standard deviation of the excess return. Thus, let R_A be the return on asset A and let R_b be the return on a benchmark (risk-free asset). The the Sharpe ratio of asset A is given by

$$S_A = \frac{E\left(R_A - R_b\right)}{\sigma}$$

where E is the expectations operator, and σ is the standard deviation of excess returns, $\sigma = (var(R_A - R_b))^{1/2}$. Note that if asset b is really risk-free then $\sigma = (var(R_A))^{1/2}$. Thus the Sharpe ratio is a risk-adjustment measure of excess return. Suppose you have two assets with the same expected excess return. If one of the assets has less volatile returns its Sharpe ratio will be higher. You can think of it as reward per unit of risk.

[20] Many of LTCM's bets were swaps, but Tiger seems to have engaged in the pure carry trade.

[21] It seems that the central banks of both governments decided to intervene to prop up the yen that summer – this was unexpected by the markets.

[22] This is referred to as forward rate bias – the forward rate over-predicts the future spot rate.

this is a test of a joint hypothesis:

1. that markets are efficient

2. the absence of a risk premium

If markets are not efficient then there is no reason to believe that UIPC will hold. In that case there are systematic errors, and agents can possibly make profitable trades. Even if markets are efficient, however, the condition may fail because of the presence of a risk premium. If the future exchange rate over-predicts the spot exchange rate it may be evidence of a risk premium. Risk averse agents are purchasing future contracts for the delivery of foreign currency because they wish to eliminate currency risk. In that case they bid up the future price of foreign currency above the spot rate. This is a case of a positive risk premium on the foreign currency. Of course, it could also go the other way. And one could find that β varies over time.

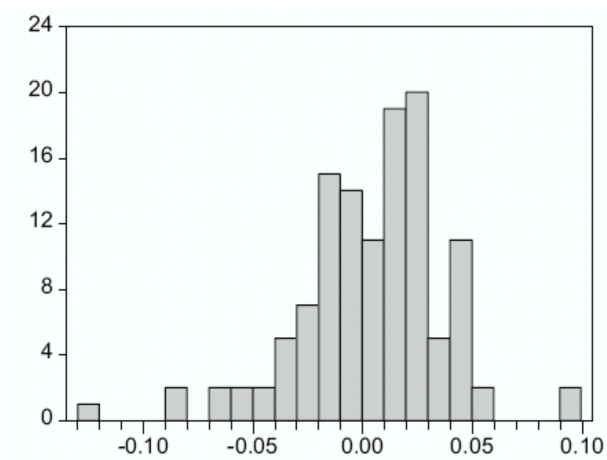

Figure 6: Realized Profits on Yen Arbitrage

If we can identify the risk premium we could then test whether this accounts for the forward rate bias. But there is no way to observe this directly. What is possible is to look at the variance of the exchange rate and plausible estimates of risk aversion and see if this could produce the risk premium that is observed. One problem is that the required risk premium is going to be time-varying: it is not constant over time.

Most studies are unable to find such a systematic relationship. Indeed, it turns out that some of the X variables are useful for explaining forecast errors. This is surprising. For example, historic forecast errors seem to be significant in explaining current forward rate forecasting errors.

This inefficiency is important. If markets were efficient then the case for intervention in currency markets would be much weaker. Of course the converse is not necessarily true – intervention could still be counter-productive. But it does open the door. So it is important to know why markets are inefficient.

What explains this inefficiency? One explanation could be that noise traders can cause the deviation from fundamental values. This could result from *insufficient* speculation. If arbitragers are liquidity constrained they may be unable to profit from short selling "overvalued" assets. Now this may seem surprising given how *large* these markets are. But many of the participants may not be interested in short-run profits. Why? Banks tend to close out their positions each night. And corporations, though they hold for long periods, may not be able to re-adjust their hedges in response to shocks due to their financial policies. They may be informed, but they may be unable to make speculative transactions when market prices stray from fundamentals.

Economists prefer to think that markets are efficient, so the typical explanation centers on the risk premium, ρ, as in expression (8). In particular, it is thought that this may vary over time. Why would risk matter? Risks arise because the future value of currencies are uncertain. If you engage in currency arbitrage, for example, you do not know what the value of the exchange rate will be when you have to convert your interest earnings back to your home currency. If currencies are volatile then such risks are larger. If agents do not like risk (if they are risk averse) then they will pay to reduce them – that means they are willing to sacrifice some income to avoid the risk – we observe that people buy insurance, after all. This means that they will demand a premium to bear a risk.

To see this, consider figure 7 where we plot income against utility. $U(y)$ is concave because the individual is risk averse: certain incomes are preferred to uncertain ones. Suppose that an individual is presented with an uncertain income stream. Specifically, with probability α income will equal y_1 and with probability $(1-\alpha)$ income will be y_2. Then average income will be $\widehat{y} = \alpha y_1 + (1-\alpha)y_2$. The utility an individual obtains from this gamble is $u(\widehat{y})$. Compare this to the utility that the individual obtains from the certain level of income $y_3 < \widehat{y}$, that is $u(y_3)$. Because $U(y)$ is concave – that is because the agent is risk averse – $u(y_3) > u(\widehat{y})$. This means that the individual prefers to have a lower level of income with certainty to the uncertain gamble. The individual would pay to reduce risk. Notice that the maximum amount that the agent would pay to reduce the risk is equal to y_4, which is referred to as the *certainty equivalent* level of income. Clearly the individual would be willing to pay any amount up to $\widehat{y} - y_4$ to be relieved of this uncertainty. Alternatively, one can think of this amount as the risk premium that the agent demands to go from the certain prospect y_4 to the uncertain prospect $\widehat{y}$.

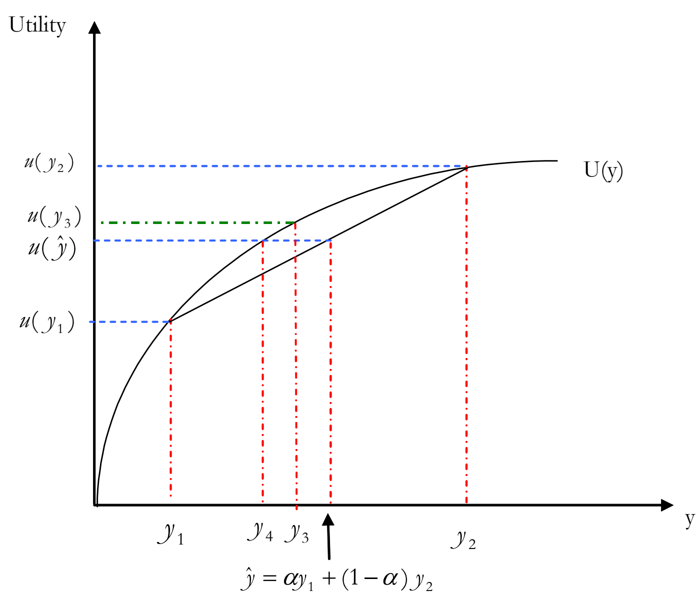

Figure 7: Certain and Uncertain Income

We can note two important points about the risk premium from figure 7.

- First, the size of the risk premium clearly depends on the extent of volatility. To see this simply make the difference between y_1 and y_2 smaller without changing $\widehat{y}$. You can see for yourself that the size of the risk premium that will be demanded will fall. So one point is that the risk premium depends on the volatility or variance of returns, in the case at hand, on the variance of the future exchange rate.

- Second, the size of the risk premium depends on the degree of risk aversion. Suppose the person was not risk averse. If agents are not risk averse then $u(y)$ will not be concave. If they are neutral to risk it will be linear. It is is easy to see that with a linear $u(y)$ the certainty equivalent level of income is equal to the average income. So the agent will pay nothing to be relieved of risk.[23] Thus, if agents are risk neutral they will not demand a risk premium. This means that all excess returns would be arbitraged away. Since we see that they are not in fact fully arbitraged – that is, we see that uncovered interest parity fails – perhaps this is due to the risk premium. And we clearly see that the degree of risk aversion effects the size of the risk premium that is demanded..

This reasoning suggests that if the volatility of income varies over time (perhaps with policy) and if attitudes towards risk vary that the risk premium that agents demand to engage in uncovered arbitrage will also vary. And this explains why there might a be a time-varying risk premium, and why you uncovered interest parity fails.

Central bank behavior may be another reason why markets are inefficient. They may be able to move exchange rates in the short run – they certainly think so, often losing large amounts of money in the process. They battle with arbitragers, and may win some times. It could be that when central banks are heavily intervening the risk premium rises. There seems to be evidence that in periods of higher central bank intervention (that is when their purchases or sales of foreign exchange are significantly above average) that the departure from *UIPC* increases. Also, it appears that the puzzle is greater in floating rate regimes than in fixed exchange rate regimes.

Some evidence for this may be that it is harder to reject *UIPC* at longer horizons. This

[23] Some people like to sky dive, for example. For them $u(y)$ would be convex; they would pay to have risks. We know that people go to Las Vegas or gamble with bookies. The house makes a profit, so people are willing to pay to absorb risk. Hence, some people are risk lovers. The puzzle are the people who drive to Las Vegas in an insured automobile. That is a bit harder to understand. One possibility is that $u(y)$ is $S-$shaped, with both convex and concave portions. Try to draw this and see if you can explain why an insured motorist purchases a lottery ticket.

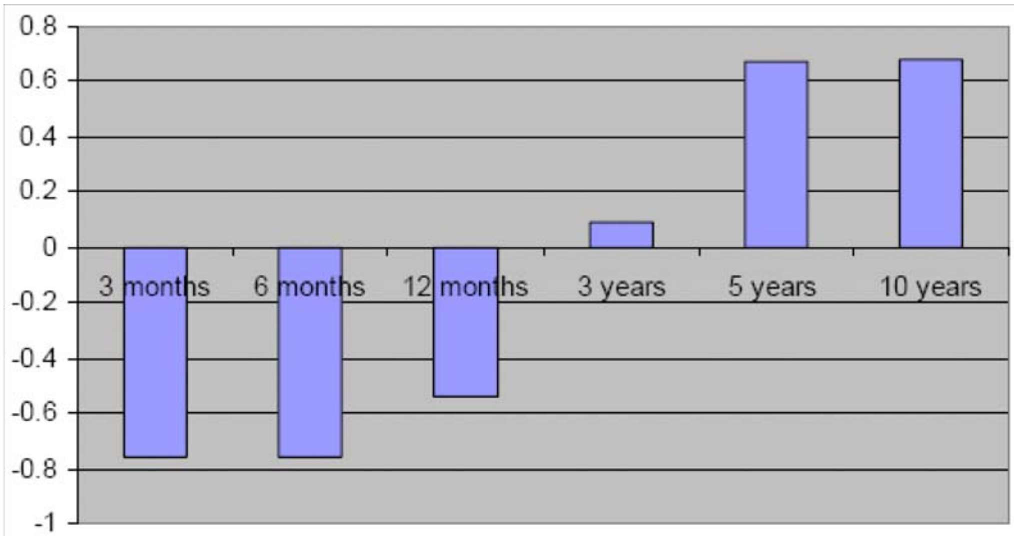

Figure 8: Estimated Beta Coefficients at Different Horizons

could mean that the risk from central bank intervention is less important at longer horizons. This is evident in figure 8 which shows the value of the estimated beta coefficient at different horizons. We can see that as the horizon lengthens, the data seems to support interest parity.

Another explanation could be peso problems (discussed below).

The problem with the risk premium hypothesis is that nobody seems able to estimate a risk premium that fits the data (that is not time-varying). Moreover, it seems that the forward premium puzzle is more apparent with regard to industrialized countries compared with emerging market economies. This makes no sense as a risk premium since the latter ought to be more risky.

2.2.1. Fama Problem

Fama showed why the risk premium is unlikely to explain the forward discount. We have seen that typically researchers find estimated values of $\beta < 1/2$ and most often negative. The problem is that $\beta < 1/2$ requires that the risk premium be more volatile than expected future exchange rates, and $\beta < 0$ implies that the covariance between the risk premium and the expected future exchange rate should be negative. This is an important result, so it is worth a bit of slogging.

To see this, first note the least squares estimate of β in a regression like (10) converges to

$$p\lim(\beta) = \frac{Cov(F_t - e_t, e_{t+1} - e_t)}{var(F_t - e_t)}. \tag{15}$$

as the sample size gets large.[24] This standard result is just telling us that the coefficient on the independent variable of interes in the regression (like expression (10) depends on the extent to which the dependent and independent co-vary relative to the variation in the independent variable. Obviously if they did not move together at all the coefficient would have a value of zero.

Now we can define the risk premium as, $\rho_t = F_t - \widehat{e}_{t+1}$, the idea being that under risk neutrality arbitrage drives $F_t \to \widehat{e}_{t+1}$ so that profits from market speculation is zero. If $F_t > \widehat{e}_{t+1}$ then investors incur a risk premium from investing in foreign currency. so

$$F_t - e_t = \widehat{e}_{t+1} - e_t + \rho_t. \tag{16}$$

We assume rational expectations. This means that the expectations error $e_{t+1} - \widehat{e}_{t+1}$ must be uncorrelated with all variables observable at t (or earlier), including F_t. Thus, our expectational errors cannot be correlated with the forward premium or discount (else we would use that information), so:

$$E_t \left\{ (F_t - e_t) \left(e_{t+1} - \widehat{e}_{t+1} \right) \right\} = 0 \tag{17}$$

which implies that under rational expecations we can write (15) as

$$p\lim(\beta) = \frac{Cov(F_t - e_t, \widehat{e}_{t+1} - e_t)}{var(F_t - e_t)}. \tag{18}$$

Now if we use (16) to substitute for $F_t - e_t$ in the numerator of (18) we have

$$\begin{aligned} Cov(F_t - e_t, \widehat{e}_{t+1} - e_t) &= Cov(\widehat{e}_{t+1} - e_t + \rho_t, \widehat{e}_{t+1} - e_t) && (19) \\ &= var(\widehat{e}_{t+1} - e_t) + Cov\left(\widehat{e}_{t+1} - e_t, \rho_t\right) && (20) \end{aligned}$$

[24]The symbol $p\lim$ is shorthand for probability limit. It means that as the sample size goes to infinity the estimator converges to this value. With enough flips of the coin, the average number of heads converges to 1/2.

where the latter follows from the simple properties of covariance.[25]

Now, note that variances are non-negative, the only way that the RHS of (19) and therefore $p\lim(\beta)$ can be negative is if $Cov\left(\widehat{e}_{t+1} - e_t, \rho_t\right) < 0$, which is Fama's second claim. In words, this means that a negative coefficient on β can occur only if expectation errors and the risk premium are negatively related. When the risk premium is high expecational errors are low and vice versa.

To see Fama's first claim, multiply both sides of (18) by $var(F_t - e_t)$ and again use (16) to substitute for $F_t - e_t$:

$$\begin{aligned} p\lim(\beta)var(F_t - e_t) &= Cov(\widehat{e}_{t+1} - e_t + \rho_t, \widehat{e}_{t+1} - e_t) \\ p\lim(\beta)var(\widehat{e}_{t+1} - e_t + \rho_t) &= var(\widehat{e}_{t+1} - e_t) + Cov\left(\widehat{e}_{t+1} - e_t, \rho_t\right) \end{aligned}$$

now multiply out the LHS,

$$\begin{aligned} & p\lim(\beta)\left(var(\widehat{e}_{t+1} - e_t) + var(\rho) + 2Cov\left(\widehat{e}_{t+1} - e_t, \rho_t\right)\right) \\ = \; & var(\widehat{e}_{t+1} - e_t) + Cov\left(\widehat{e}_{t+1} - e_t, \rho_t\right) \end{aligned}$$

so

$$p\lim(\beta) = \frac{var(\widehat{e}_{t+1} - e_t) + Cov\left(\widehat{e}_{t+1} - e_t, \rho_t\right)}{var(\widehat{e}_{t+1} - e_t) + var(\rho) + 2Cov\left(\widehat{e}_{t+1} - e_t, \rho_t\right)}$$

so that if $p\lim(\beta) < 1/2$, it follows that:[26]

$$\frac{1}{2}\left\{var\left(\widehat{e}_{t+1} - e_t\right) + var(\rho_t)\right\} > var\left(\widehat{e}_{t+1} - e_t\right) \tag{21}$$

[25] To see, let x and y be any two random variables. Then we can write the expression in the text simply as:

$$Cov(x + y, y) = Cov(x, x) + Cov(x, y)$$

But $Cov(x, x) = var(x)$, so

$$Cov(x + y, y) = var(x) + Cov(x, y)$$

which is the expression we sought to show. More simply, note that

$$\begin{aligned} Cov(x + y, y) &= E[(x + y)x \\ &= E[x^2 + xy] = E[x^2] + E[xy]. \end{aligned}$$

[26] This follows because obviously $\frac{Cov(\widehat{e}_{t+1} - e_t, \rho_t)}{2Cov(\widehat{e}_{t+1} - e_t, \rho_t)} = \frac{1}{2}$.

and thus $var\left(\widehat{e}_{t+1}-e_t\right)+var(\rho_t)>2var\left(\widehat{e}_{t+1}-e_t\right)$ so that

$$var(\rho_t)>var\left(\widehat{e}_{t+1}-e_t\right) \tag{22}$$

which is Fama's first claim. This says that the variance of the risk premium must be greater than the variance of the expecation errors. So we need a lot of risk and it should be volatile, it seems.

Expression (22) seems quite unlikely to hold and thus would appear to be a serious argument against the risk premium assumption. One should note, however, that (22) can hold either if the LHS is large or the RHS is very small. Indeed, if expected changes in exchange rates are small then (22) is not so puzzling. And if exchange rates are a random walk, then $var\left(\widehat{e}_{t+1}-e_t\right)=0$. Thus the surprise may not be that the variance of the risk premium is so large, but rather that the expected change in the exchange rate is so small.

2.2.2. Other issues

Frankel and Poonawala looked at the forward discount puzzle in emerging economies. They found that while the bias in the forward discount as a predictor of the future change in the spot exchange rate is present among emerging market currencies and advanced country currencies alike, the bias is less severe in the former case than in the latter. Unlike major currencies, which generally show a coefficient significantly less than zero, suggesting that the forward rate actually points in the wrong direction, the coefficient for emerging market currencies is on average slightly above zero, and even when negative is rarely significantly less than zero.

One implication for traders is that the "yen carry trade" and "dollar carry trade" on average may not be as profitable when the strategy is to go long in emerging market currencies as when it is to go long in major currencies. An implication for international finance theorists, in light of the intuitively high riskiness of emerging currencies, is that the source of forward discount bias may not lie entirely in the exchange risk premium.

It could be that bid-ask spreads are such that one cannot make money taking advantage of it. We have seen that carry trade is profitable, but it could be that exploiting these small gains would require trading of sufficient volume to make the anomaly go away. Why would this be the case? Typically, we have ignored bid-ask spreads as they are small. The problem is that although they are small, so are the returns on the carry trade. When we take these into account the amount we need to trade to make money on the carry trade is large. A speculator who be one pound on an equally-weighted portfolio of carry-trade strategies (across the USA, Canada, Belgium, France, Germany, Japan, Netherlands, Switzerland and the euro) from 1976 to 2005 would earn an monthly payoff of 0.0025 pounds. To earn an average annual payoff of 1 million pounds would require a bet of 33.33 million pounds per month.

Now suppose that you actually try to trade this much. Is there an effect? It turns out that the price in the foreign exchange market tends to be affected by the order flow. That is, if trading is really high the price rises – this is called price pressure. Price pressure occurs when the price at which an investor can buy or sell depends on the size of the transaction. The magnitudes are not large, but they eat into the returns. So what might you do? You could break up the sales into smaller chunks. But that would require more trades which adds tiny transactions costs. So it could be that there are average excess returns, but there is no money left on the table if you try to earn them.

Peso Problems A peso problem arises when the underlying variables that affect some asset price are subject to very different discrete possibilities. In this case expectations are likely to turn out incorrect even when made properly. Think of rain or shine. In one case you need an umbrella in the other you do not. If there is 70% chance of rain you take the umbrella. But if it does not rain your guess was wrong. In such cases you will likely make many mistakes even if your prediction mechanism is pretty good.

The term peso problem is often attributed to Milton Friedman in comments he made about the Mexican peso market of the early 1970s. At that time, the exchange rate between the U.S. dollar and Mexican peso was fixed, as it had been since 1954. At the same time, the interest rate on Mexican bank deposits exceeded the interest rate on comparable U.S. bank deposits. This situation might seem like a flaw in the financial markets, since investors could borrow at the low interest rate in the United States, convert dollars into pesos, deposit the money in Mexico and earn a higher interest rate, then convert the proceeds back into dollars at the same exchange rate and pay off their borrowings, making a tidy profit. Friedman noted that the interest rate differential between Mexico and the United States must have reflected financial market probabilities of bad (or good!) things happening, concerns that the peso would be devalued. Otherwise, the interest-rate differential would soon disappear as investors increasingly took advantage of it. In August 1976, those concerns were justified when the peso was allowed to float against the dollar and its value fell 46 percent. The difference in return on comparable U.S. and Mexican assets—which looked like an anomaly to analysts who thought the exchange rate would remain fixed because it had been fixed for 20 years—could be explained once investors' recognition of the possibility of a large drop in the value of the peso was factored in.

Example 1 *Suppose the spot exchange rate is 20 cents to the peso, and this has been fixed for a while. Suppose that investors believe that there is a 95% chance it will remain at 20 cents, and a 5% chance it will fall to 10 cents per peso. What is the expected value of the peso?*

$$EV = .95 * 20c + .05 * 10c = 19.5c$$

As long as the peso is not devalued what is the forecast error? It is $\varepsilon_t = 20c - 19.5c = 0.5c$. *Notice that this forecast error will persist for every period until the devaluation. A casual observer may conclude that the forecasters are irrational since they are consistently wrong. But this is a result of the zero-one possibility here.*

Example 2 *Suppose that the current interest rate in the US is* 5%. *Then* UIPC *implies that*

$$\begin{aligned} i^* &= \frac{e_t}{\widehat{e}_{t+1}}(1+i) - 1 \\ &= \frac{.20}{.195}(1.05) - 1 = .076923 \end{aligned}$$

The currency markets will signal a depreciation of the peso, but each period that it does not *happen will imply the market is inefficient.*

2.2.3. Comparison to Real Interest Parity

Uncovered interest parity suggest that interest differentials (nominal) are equal to expected changes in the currency. What about real interest differentials? The real interest rate, r, is equal to the nominal rate minus expected inflation, $r = i - \pi^e$. So $r - r^* = (i - i^*) - (\pi^e - \pi^{*e})$. If PPP holds, then $\Delta s^e = \pi^e - \pi^{*e}$, so $r - r^* = (i - i^*) - \Delta s^e$. But uncovered interest parity implies that $(i - i^*) - \Delta s^e = 0$, so that real interest differentials are equal to zero. Real interest differentials are arbitraged away. But PPP is a restrictive assumption. What happens in general?

The assumption of PPP is equivalent to assuming that the real exchange rate is constant. Yet, we know it is not. So how is the theory modified? Recall that the real exchange rate is defined as $Q_t = \frac{e_t P_t^E}{P_t^{US}}$. We are interested in an expression for the *expected* growth rate of the real exchange rate, $\frac{Q_t^e - Q_{t-1}}{Q_{t-1}}$. Suppose that inflation was expected to equal in the US and euroland. Then clearly we would have $\frac{Q_t^e - Q_{t-1}}{Q_{t-1}} = \frac{e_t^e - e_{t-1}}{e_{t-1}}$. Of course expected inflation rates are not equal, however, so how is the expression altered? Suppose that the exchange rate was not expected to change. Then clearly we would have $\frac{Q_t^e - Q_{t-1}}{Q_{t-1}} = \pi_E^e - \pi_{US}^e$. If US inflation is higher than in euroland the real exchange rate depreciates. Put these two factors together and it is clear that:

$$\frac{Q_t^e - Q_{t-1}}{Q_{t-1}} = \frac{e_t^e - e_{t-1}}{e_{t-1}} - (\pi_{US}^e - \pi_E^e). \tag{23}$$

The expression is intuitive.[27]

[27] To prove this take logs of the expression for the real exchange rate expression:

$$\log Q_t = \log e_t + \log P_t^E - \log P_t^{US}$$

Now differentiate both sides with respect to time and we obtain:

$$\left(\frac{1}{Q_t}\right)\frac{dQ_t}{dt} = \left(\frac{1}{e_t}\right)\frac{de_t}{dt} + \left(\frac{1}{P_t^E}\right)\frac{dP_t^E}{dt} - \left(\frac{1}{P_t^{US}}\right)\frac{dP_t^{US}}{dt}.$$

But these expressions are just the (continuous time) growth rates of the real and nominal exchange rate and of the price levels in euroland and the US, respectively. So as the period shrinks, we obtain the expression in the text.

Now recall the interest parity condition, $\frac{e_t^e - e_{t-1}}{e_{t-1}} = i_{US} - i_E$. Using this to replace the expected rate of depreciation in (23), we obtain:

$$i_{US} - i_E = \frac{Q_t^e - Q_{t-1}}{Q_{t-1}} + (\pi_{US}^e - \pi_E^e) \tag{24}$$

which implies that interest differentials depend on expected movements in the real exchange rate in addition to differences in expected inflation. When the real exchange rate is not expected to change we have the same expression as before. But in general, nominal interest differentials are explained by movements in the real exchange rate as well. Of course you can think of the latter as capturing all the reasons for exchange rate movements other than differential price levels.

This now lets us derive an expression for *real interest parity.* While nominal returns are those that are the actual components of exchanges, it is real returns, or rather *expected* real returns that govern decisionmaking. That is, you may earn a nominal interest rate of 10% from an asset, but whether you choose to hold it or not depends on how the real return relates to the opportunity cost of the funds. So an investor thinking of where to hold her wealth must consider the expected real returns.

We know that nominal interest differentials are related to expected changes in the real exchange rate and expected inflation. We obtain the real interest parity condition by first using the Fisher equation, $i_t = r_t + \pi^e$. It follows that the expected real rate of interest in the US, $r_{US,t}^e = i_{US,t} - \pi_{US,t}^e$, and likewise for Euroland, $r_{E,t}^e = i_{E,t} - \pi_{E,t}^e$. Using these in expression (24) yields the real interest parity condition:

$$r_{US,t}^e - r_{E,t}^e = \frac{Q_t^e - Q_{t-1}}{Q_{t-1}} \tag{25}$$

which says that expected real interest rate differentials are equal to expected changes in the real exchange rate.

Why should (25) hold? Suppose that people expect the real exchange rate to appreciate, say because Euroland productivity growth in tradables is expected to be higher than in Euroland non-tradables, and also higher than in the US. Thus people expect the real value of the dollar to depreciate relative to the euro. To compensate for this the real return on dollar assets must exceed those of Euroland assets.

Does this mean that there are profit opportunities that are not being arbitraged away? No. The differences in the real rates do not reflect different returns on the *same asset*. It reflects different returns on two bundles of goods. The absence of arbitrage opportunities is guaranteed by interest parity, since any investor that compares relative returns has a unique consumption basket. When I compare the rate of return on holding dollars or euros, the real return is computed by subtracting my expected rate of inflation, whatever consumption basket is relevant for me. But the expected real differential on the left-hand side of (25) is comparing two expected inflation rates that reflect two *different* consumption baskets. Notice that if all agents were identical PPP would hold and we would not have real interest differentials. But because people in different countries consume different baskets of goods, there is no way for them to arbitrage away any difference.

Why is this interesting? Recall when we spoke of the Feldstein-Horioka puzzle we commented that a direct way to test for capital market integration would be to look at whether excess returns were arbitraged away. So that would suggest looking at real interest differentials in different countries. If capital markets are integrated these should go to zero. But as we have just seen, this is only true if PPP holds. If it does not hold, then real interest differentials should be expected to persist. Suppose, for example, that US savings was so low that people expected the real exchange rate to appreciate in the future.[28] Then real returns on US assets would have to exceed those in the rest of the world, according to (25).

3. Fixed, Flexible, and Managed Exchange Rates

Finally, let us distinguish among exchange rate regimes. There are two polar cases to consider – fixed versus flexible – along with various gradations. Usually in economics fixed prices are considered problematic; flexible prices is the norm and ideal. In the exchange

[28] Why do they expect Q to appreciate? We have to pay back the debts we have incurred, so current account deficits must become surpluses. So the relative price of our goods must fall relative to the rest of the world to allow us to export more and import less.

market, however, it is not so simple. There is no simple correlation between free market principles and the exchange rate regime. The range of opinion on this issue can perhaps be best illustrated by the contrasting views of Robert Mundell and Milton Friedman. Mundell believes that any departure from fixed exchange rates is an abomination, because exchange rate changes by definition imply reneging on the promised value of a currency.[29] Friedman, on the other hand, believes that fixing any price is wrong, and the exchange market is no different.[30] Indeed, Friedman makes an interesting analogy with regard to flexible exchange rates, likening them to the argument for daylight savings time. First, he notes that exchange rates are potentially more flexible than internal prices:

> If internal prices were as flexible as exchange rates, it would make little economic difference whether adjustments were brought about by changes in exchange rates or by equivalent changes in internal prices. But this condition is clearly not fulfilled. The exchange rate is potentially flexible...At least in the modern world, internal prices are highly inflexible ("The Case for Flexible Exchange Rates, in *Essays in Positive Economics*).

Friedman went on to offer a famous analogy:

> The argument for flexible exchange rates is, strange to say, very nearly identical with the argument for daylight savings time. Isn't it absurd to change the clock in summer when exactly the same result could be achieved by having each individual change his habits? All that is required is that everyone decide to come to his office an hour earlier, have lunch an hour earlier, etc. But obviously it is much simpler to change the clock that guides all than to have each individual separately change his pattern of reaction to the clock, even though all want to do so. The situation

[29] Then it is not surprising that Mundell is a strong advocate of the gold standard.

[30] Neil Wallace disputes this last contention. Unlike goods markets, the foreign exchange market involves trade in fiat currencies. This is fundamentally different than trade in goods – there are no fundamentals driving market clearing. The supplies of fiat currencies must be controlled for them to have value, but the level is irrelevant. If there is more money each unit has less value, but money has the same value. Hence, it is legal restriction that keeps people using pesos or dollar. In the absence of this, or in the absence of expected government intervention in the exchange market, floating exchange rates are indeterminate.

is exactly the same in the exchange market. It is far simpler to allow one price to change, namely, the price of foreign exchange, than to rely upon changes in the multitude of prices that together constitute the internal price structure.

We will return to this discussion later when we compare the advantages and disadvantages of various exchange rate regimes. For now we want to see how the exchange rate is determined. It is useful to begin with a partial equilibrium analysis of the market for foreign exchange to fix ideas.

Let the demand for foreign exchange be given by $D(e, Y, \Phi)$, where Y is domestic national income, and Φ is meant to capture all other factors that affect the demand for foreign exchange. The demand for foreign exchange depends negatively on the exchange rate, because a higher value of e means that it takes more dollars to purchase a unit of foreign exchange. D depends positively on Y because higher income means more imports, and this requires the purchase of foreign exchange. The supply of foreign exchange is given by $S(e, Y^*, \Psi)$, where Y^* is income in the foreign country, and Ψ represents shift variables for supply.[31] Clearly, $S_e > 0$, since a higher exchange rate means that exports are cheaper. Similarly, higher foreign income also increases the supply of foreign exchange because it increases our exports.

The market-clearing exchange rate is given by

$$D(\widetilde{e}, Y, \Phi) = S(\widetilde{e}, Y^*, \Psi) \tag{26}$$

and in figure 9 by $\widetilde{e}$. If the exchange rate regime was pure floating, then the exchange rate would be determined by this market clearing condition (26). This is the regime of purely flexible exchange rates.[32] Shocks to the demand or supply of foreign exchange would result in movements of the exchange rate. For example, if Y increased, this would cause D to shift to the right, and $\widetilde{e}$ would increase.

In a regime of purely flexible exchange rates central banks do not intervene in the exchange market. Hence, the only flows are those created through the current account and the capital

[31] Notice that capital flows can affect the supply and demand for foreign exchange, and the variables that affect such flows are subsumed in Φ and Ψ. Of particular importance in this regard are interest rate differentials.

[32] The essential point is that central banks do not intervene in the market for foreign exchange.

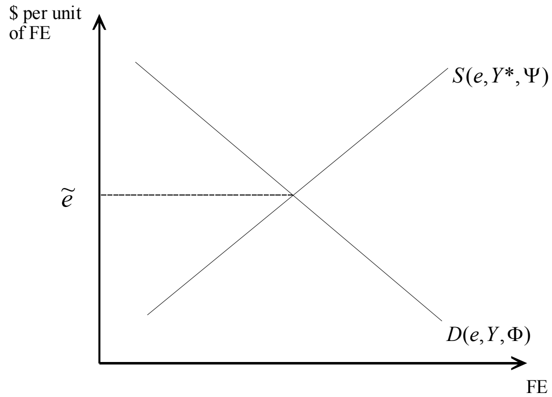

Figure 9: The Supply and Demand for Foreign Exchange

account; that is, by private flows.[33] Thus a regime of pure floating can also be characterized by setting the official settlements balance (ΔIR) equal to zero, so that

$$CA_t + KO_t = 0. \tag{27}$$

We will have occasion to return to this expression.

Fixed exchange rates occur when the central bank commits to a particular exchange value for domestic currency, and stands ready to buy or sell foreign exchange to maintain that rate.[34] Hence, with fixed exchange rates any excess demand for foreign exchange is met by sales of foreign exchange by the central bank:

$$D_t(\overline{e}, Y, \Phi) - S_t(\overline{e}, Y^*, \Psi) = -\Delta IR_t \tag{28}$$

where $\overline{e}$ is the fixed exchange rate, and ΔIR is the change in foreign reserves held at the central bank. Thus if at $\overline{e}$ there is an excess demand for foreign exchange, the central bank sells its reserves of foreign exchange to prevent the exchange rate from increasing (the currency from depreciating), hence its supplies of international reserves are going down ($\Delta IR < 0$).

[33]Technically even with flexible rates governments may purchase foreign currency to affect foreign transacations – for example, to pay for expenses of embassies. The key point, however, is that the central bank does not alter its holding of foreign exchange.

[34]The gold standard is one example of a fixed exchange rate, and we shall examine it in more detail. Notice that under such a regime $\overline{e}$ would be the dollar price of gold, and foreign exchange would refer to gold. A country on the gold standard stands ready to convert domestic currency into gold at the fixed parity, $\overline{e}$. The gold standard is thus a special case of a fixed exchange rate regime. Of course, in a pure gold world there is just a single monetary unit, gold. But under a gold standard there is a domestic currency that trades to gold at a fixed rate supported by the central bank.

While the central bank can fix the exchange rate in this manner by passively responding to the excess demand on the LHS of (28), this is not the only way. The central bank could also peg the exchange rate by using monetary or fiscal policy to affect the market clearing rate. By raising interest rates the central bank could attract capital flows and thereby increase the demand for domestic currency. We will discuss this in more detail later.

Suppose, however, that policy has not succeeded in moving the market clearing rate to equal the fixed exchange rate. Instead, suppose that $\overline{e} < \widetilde{e}$; the dollar is over-valued and there is an excess demand for foreign exchange. We can see from (28) that reserve sales will offset this excess demand. In this setting the reserve flow appears to be passive, or *induced*. When capital flows were much restricted (in the early Bretton Woods period) this was referred to as an induced capital flow. This led to the practice whereby the changes in the official reserves settlement balance was treated as caused by the current and capital account. Today, central banks do not act so passively, so this terminology is no longer common.

It is important to note an important asymmetry. When an exchange rate deviates from the market clearing rate one currency is overvalued while another is undervalued. If $\overline{e} < \widetilde{e}$ the dollar is over-valued and the Fed must sell foreign exchange. At the same time, the DM must be undervalued, so the Bundesbank would be accumulating dollars.[35] Neither country's market for foreign exchange is in equilibrium, but the positions are not symmetric. The reason is that reserves are declining in the US but accumulating in Germany. The Bundesbank is selling DM and accumulating dollars. It can do this forever, since the Bundesbank can print as many DM as needed. The only cost is that of building warehouses to hold dollars. But the Fed has only a finite quantity of foreign reserves. If reserves were to run out the Fed would have to let e adjust.

Notice that we might have $\overline{e} < \widetilde{e}$ for two different reasons. First, it could be just a temporary deviation due, for example, to a transitory decline in Y^*. Transitory fluctuations

[35] The countries with the largest foreign reserves are Japan ($844 billion, September 2005), China (Mainland) $769 billion, Republic of China (Taiwan) $254 billion, Hong Kong $123 billion, September 2005), the Republic of Korea ($207 billion, September 2005), Russia ($160 billion, September 2005), India ($143 billion, September 2005), Singapore($116 billion, August 2005), Germany ($100 billion, September 2005) and Malaysia ($80 billion, September 2005).

in the excess demand for foreign exchange would not have any persistent effect on reserves. Alternatively, exchange rates could be out of equilibrium due to some permanent change. This means that the reserve changes will continue until some adjustment takes place. Note, however, that the burden of adjustment is much greater on the deficit country.

A country that is experiencing a persistent balance of payments deficit – we use that terminology again to refer to the status of ΔIR – will presumably adjust before reserves reach zero. After all, if the exchange rate is going to be changed there is no reason to let all reserves be wiped out. But what the exact level of that threshold happens to be is unknown until it happens. Moreover, central banks may wish to keep the actual level of reserves secret precisely when under such pressure to keep speculators from turning on the currency.

3.1. Reserve Accumulation under Fixed Rates

One might ask if there is any effect of sterilizing reserves. There is an effect on the money supply. To see this it is useful to think of a simple version of the central bank's balance sheet:

assets	*liabilities*
Foreign reserves (IR)	Currency in circulation
Domestic securities (DS)	Bank reserves

Because assets must equal liabilities, we can note that the assets of the central bank, $IR + DS$ sum to equal to the monetary base (MB) (or high-powered money). Note that the money stock is equal to the product of the monetary base and the money multiplier, μ:

$$M = \mu MB = \mu(IR + DS) \tag{29}$$

Now let us consider what happens when the Federal Reserve purchases foreign exchange with dollars. There are four cases to consider:

1. purchase from home-country banks: in this case alongside the increase in IR is an increase in bank reserves.

2. purchase from home-country non-bank residents: in this case, residents would receive payment in the form of currency in circulation.

3. purchase from foreign banks or central banks : in these cases currency in circulation rises immediately if the payment is made in cash.

4. purchase from foreign banks or central banks via changes in the foreign bank's deposit at the Fed. In this case, once the bank uses this deposit to purchase some interest-bearing security from a domestic bank, bank reserves will rise.

The key point, in any case, is that reserve transactions must result in simultaneous changes in the money base. This makes the stock of money endogenous. If foreign reserves are increasing so is the domestic money stock, and vice versa.

Sterilization There is one way that the domestic money stock can be insulated from reserve changes. Suppose that at the same time the Fed purchases foreign exchange it also sells domestic securities; that is, it engages in an open market operation. The latter transaction will decrease the money stock, and total central bank assets will be unaffected. This action is called *sterilization* because the domestic economy is insulated from the foreign reserve transaction. This can be costly, however, as the country must borrow from domestic citizens at a higher rate that it earns on its reserves. So there is a fiscal cost.[36]

In practice, however, it turns out that sterilization is very difficult to achieve. To sterilize, the Central Bank must change the stock of domestic securities to maintain unchanged the stock of high-powered money. That is,

$$\Delta H = \Delta IR + \Delta DS = 0. \tag{30}$$

Thus, if the Central Bank is purchasing foreign exchange, so that reserves are increasing, it must simultaneously sell domestic securities; i.e., $\Delta IR = -\Delta DS$. Notice that to persist in sterilized intervention requires large stocks of both foreign reserves and domestic securities. The reason for the former (that of a "deficit" country) is obvious. But consider the case of a surplus country. It is accumulating reserves. Presumably it can do this forever. But to

[36] Notice that a country like China, which has capital controls, can reduce this cost because it limits the investment options of its citizens. But it cannot eliminate them entirely.

sterilize such a flow it must be selling domestic securities. But this requires that the Central Bank have a very large stock of debt to sell. This condition is unlikely to be satisfied in most economies.

4. Gold Standard and the Specie-Flow Mechanism

The gold standard is an important historic example of a fixed exchange rate regime. To some it is the ideal system because it takes central banks out of the exchange rate business. It is also seen as a natural system, as opposed to one that derives from government behavior. This is really a simplification of how the gold standard really worked. And it ignores the fact that the gold standard only really worked under peculiar historic circumstances.

It is useful nonetheless to examine Hume's specie-flow mechanism. This is really the first general equilibrium adjustment model in economics. It is a powerful model, and it is still useful, though it is based on simplified assumptions which produce the sharp predictions.[37]

Consider a world where prices are flexible and where all transactions take place with gold coins. Also assume that there is a fixed supply of gold in the world. These coins are minted at a fixed parity in each country. There are no banks and no capital flows.[38] Whenever goods are exported a merchant receives payment in gold. Not wanting gold, the merchant takes this to the mint and receives gold coins at the fixed parity. To purchase imports a merchant pays with gold. To get the gold the merchant takes coins to the mint and sells them for gold at the fixed parity, which is then used to pay for the imports. The recipient in the foreign country takes the gold to the mint and obtains coins.

Now consider the case of a country with a trade surplus. This means that more gold is coming in to the country than is leaving. Hence, the supply of gold coins in the domestic

[37] Hume wrote: "Suppose four-fifths of all the money in Great Britain to be annihilated in one night...what would be the consequence? Must not the price of all labour and commodities sink in proportion, and everything be sold as cheap as they were in those ages? What nation could then dispute with us in any foreign market, or pretend to navigate or to sell manufactures at the same price, which to us would afford sufficient profit? In how little time, therefore, must this bring back the money which we had lost, and raise us to the level of all the neighbouring nations? Where, after we have arrived, we immediately lose the advantage of the cheapness of labour and commodities; and the farther flowing in of money is stopped by our fullness and repletion." Hume, *Of Money*.

[38] Or you can think of banks with 100% backing of notes by gold.

economy is increasing. With more circulating medium (specie) the price level will increase. This follows from the assumption of price flexibility and full employment.[39] The increase in domestic coinage increases aggregate demand and pushes up prices. But this reduces the competitiveness of exports and increases the attractiveness of imports. Hence, the flow of gold will be reversed until prices return to their equilibrium levels. This is Hume's specie-flow mechanism, a complete explanation of the adjustment of price levels to shifts in the money supply across countries.

Notice that extending this model to a world with banks and paper money is not all that difficult. Assume that paper currency exists but that central banks stand ready to convert paper currency into gold at the fixed rate, as is the case in the foreign country. Suppose that Great Britain runs a trade surplus with France. That means that British exporters are accumulating franc notes. Not needing that many notes, they present the excess to the Bank of France for exchange into gold.[40] They then take the gold to the Bank of England and get more sterling notes. The domestic money supply rises, and the same adjustment as before, takes place. The systems works pretty much like it does with coin. Gold flows cause relative prices to change in Britain and France in a manner that reverses the gold flow.

Notice that because gold flows offset trade imbalances through adjustments in price levels, movements in prices tended to be cyclical rather than trending. Without growth in the world supply of gold inflation could not be sustained. Of course discoveries of gold, such as the Spanish discoveries in the Americas, did lead to increases in general prices, but the world trend in prices depended on the balance between the growth in production of goods and the growth in the production of gold.

A Model To better understand the gold standard it is good to first ask how the gold standard works in a single country. Because we have a gold standard, the dollar price of gold, P^G, is given. In the short-run the gold supply is fixed at G^0. We also have a demand function

[39] Though Hume did not argue that this would happen automatically; rather it would take some time so that at first output would increase before prices.

[40] This probably involves the intermediation of British banks which transfer to their correspondents in France.

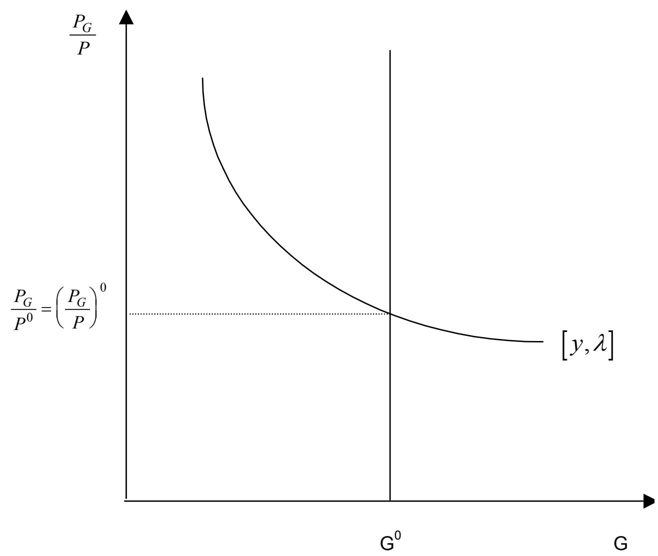

Figure 10: Price-level determination with a given Gold Stock

for gold, as in figure 10, which is drawn with a negative slope.

What determines the demand for gold? Gold is used for two purposes: non-monetary, such as jewelry, dentistry, etc., and monetary. The non-monetary demand for gold will be a decreasing function of its relative price, $\frac{P^G}{P}$. It will also depend on real income, y. If economic activity increases so should the non-monetary demand for gold.

Monetary demand for gold depends on its use for transactions. Suppose that gold is used to back the currency. Then if G_M is the quantity of gold used as reserves, and M is the stock of money, then $\lambda = \frac{P^G G_M}{M}$ is the reserve ratio. If gold were used as money, then λ would equal unity. In general it satisfies $0 < \lambda \leq 1$.

This implies that the quantity of monetary gold demanded depends on the public's demand for dollars. Thus,

$$G_M = \frac{\lambda M}{P^G} \tag{31}$$

But this depends, according to standard theory, on real income and the price level. Hence, $M = PL(y)$, where $L' > 0$. Hence, using this in 31 we obtain

$$G_M = \frac{\lambda P L(y)}{P^G} \tag{32}$$

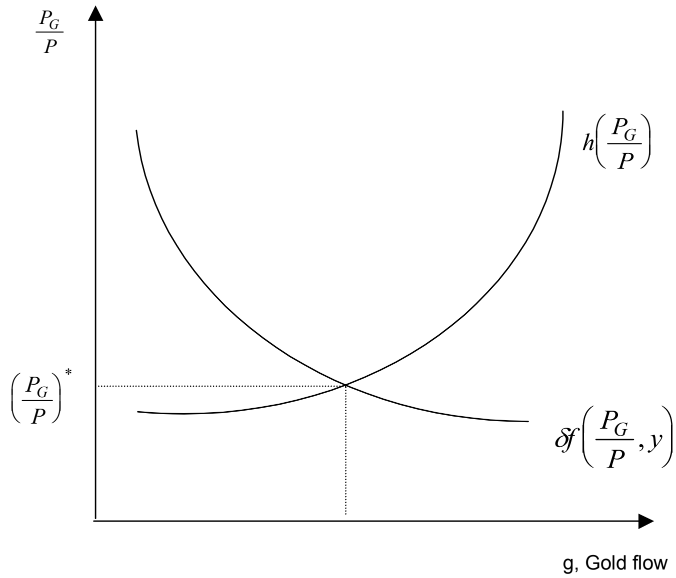

Figure 11: Flow supply and demand for gold.

which implies that the monetary demand for gold depends positively on real income and λ, and negatively on the relative price of gold in terms of goods.

Because both the monetary and non-monetary demand for gold depends negatively on the relative price, and positively on y, λ we obtain the function as in figure 10. Notice that, because the price of gold is fixed, the supply of gold determines the price level. If the stock of gold increased, or real income decreased, this would increase the price level.

We have examined the consequences of a given stock of gold, but what determines the stock? Consider figure 11, which shows the flow supply and demand for gold. The flow supply depends positively on the relative price of gold, because it is more profitable to produce gold when its price is higher.

Subtractions from the gold stock depend on wastage. We denote by δ the depreciation of the gold stock in non-monetary use, and $f\left(\frac{P_G}{P}, y\right)$ is the non-monetary demand for gold. We are assuming that money held as reserves does not depreciate. At the relative price $\left(\frac{P_G}{P}\right)^*$ the flow supply and demand are equal, so the stock of gold is not changing. If the relative price were higher than this then the flow supply would exceed demand and the stock would grow, and vice versa.

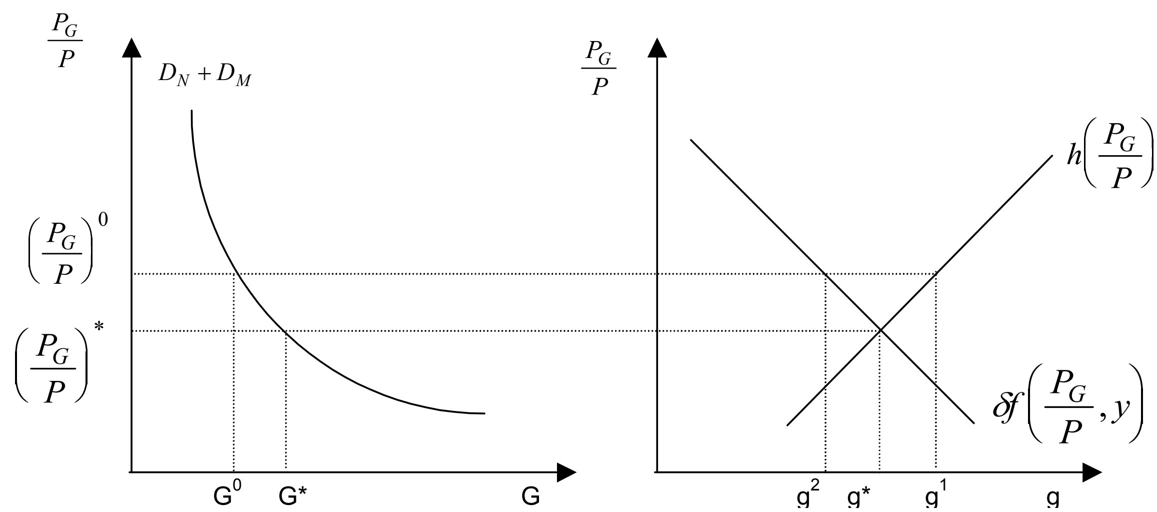

Figure 12: Gold Standard Model

We now put the two analyses together. Consider figure 12 If the initial gold stock is G^0, then the price level is $\left(\frac{P_g}{P}\right)^0$. This is the short-run equilibrium price level. But at this price the flow supply of gold is greater than its demand. This causes the stock to rise. Once the stock of gold increases to G^*, however, flow supply and demand are equal so that we are in long-run equilibrium.

Suppose that a new gold field was discovered, increasing the flow supply of gold. Suppose we were initially in long-run equilibrium. The discovery shifts $h(\cdot)$ to the right, causing the stock of gold to increase at the initial relative price of gold. As the stock increases the relative price of gold falls until the flow supply equals demand.

What about a shift in λ? Suppose that the Central Bank reduces λ. This would cause the demand function for monetary gold to fall, shifting the demand function to the left. At the given stock of gold, this would cause the relative price of gold to fall. But this would cause the flow supply to fall below demand, causing the stock of gold to fall. Eventually, the stock of gold falls sufficiently so that the relative price of gold returns to its initial level. We should note, however, that if the CB starts to use λ too much as a policy instrument we cease to have the gold standard.

Now how do we modify the model for the open economy? This turns out to be quite easy.

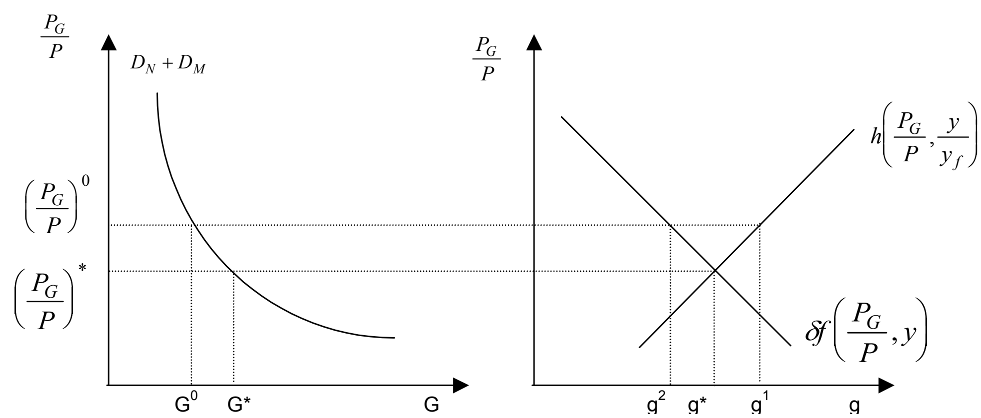

Figure 13: Gold Standard in the Open Economy

All we do is replace the flow supply function. Instead of depending on production, we now have it depend international trade in goods and services. If exports exceed imports the supply of gold increases, and vice versa. But what will the trade balance depend on? Clearly, this will depend on the relative price of gold. If the domestic price level increases then imports will increase. It will also depend on the domestic and foreign income, $\frac{y}{y_f}$. So we can now write the flow supply function as $h\left(\frac{P_G}{P}, \frac{y}{y_f}\right)$, with $h_1 > 0$, and $h_2 < 0$. we know have figure 13. The key difference now, however, is that the flow supply of gold will adjust much faster. It no longer depends on production, but on the flow of gold due to trade.

The key feature of the gold standard model is that the price level depends on the stock of gold. We can see here the specie-flow mechanism at work. If the stock of gold increased due to some discovery the relative price of gold decreases, which is the same thing as saying that our price level rises relative to foreigners (who are also tied to gold). This means that the flow supply will be less than demand, causing G to fall. The decrease in G causes our prices to fall until we return to the initial equilibrium.

Notice that under the gold standard changes in the demand for money (or changes in income) can have short-run effects, but no long-run effects on inflation. Suppose money demand increases. This would shift the stock demand for gold to the right, causing the relative price of gold to rise. This would imply that we are more competitive, and the gold

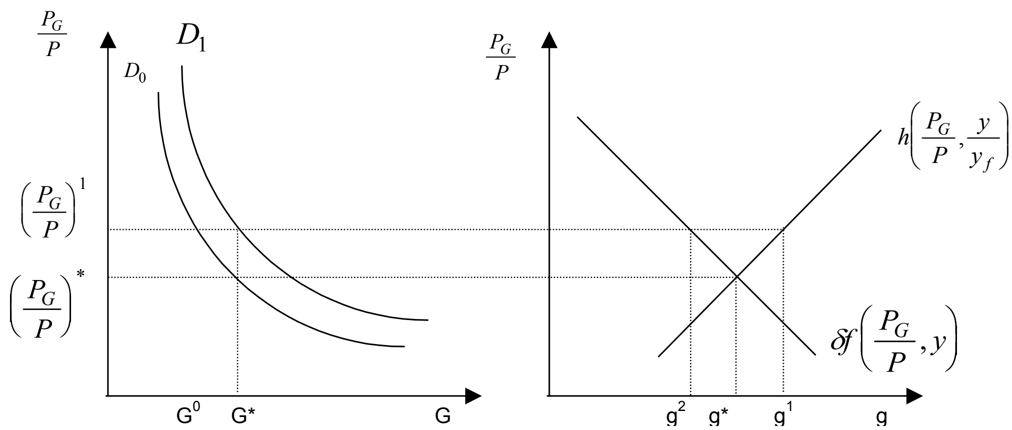

Figure 14: An increase in money demand

inflow would exceed its depreciation. The ensuing increase in the stock of gold would cause our prices to rise until we returned to the initial equilibrium relative price. This is analyzed in figure 14.

Our model of the gold standard works quite well to explain the automatic adjustment of the system. It can – unlike Hume's version – account for capital flows as well, since we have paper money. We would only need to make the flow supply of gold depend on the interest differential in addition to the relative price of gold and relative income, $h\left(\frac{P_G}{P}, \frac{y}{y_f}, \frac{r}{r_f}\right)$, where $h_3 > 0$ because a positive interest differential would cause capital to flow into the economy. This would require foreigners to purchase domestic currency with their currency, which would then be traded for gold, increasing the supply at home.

The major way in which this model departs from the actual operation of the gold standard is that gold flows were not, in fact, as large as the model suggests. The model suggests that gold flows are on the order of magnitude of the trade balance (or the trade balance plus the capital account if we consider capital flows). In fact, however, gold flows were much smaller than this level. How could this be?

The explanation for the missing gold flows is monetary policy. Although the ideal system is passive, in fact monetary authorities could intervene to speed up monetary adjustment. Rather than wait for gold flows to move the price level, a central bank could undertake policy to move the money supply in anticipation of the gold flows. Suppose that our price level is too high: at the initial relative price of gold we are running a trade deficit. Over time gold would flow out of the country, reducing our money supply, and restoring external balance. But this could take time. To speed this up, and to avoid the loss of gold, the central bank could act to tighten monetary policy *in anticipation* of the gold outflow. This could occur via an increase in the discount rate, for example.

When the central bank was tightening in anticipation of an outflow, and vice versa, this was referred to as "playing by the rules of the game." The phrase is Keynes's and thus postdates the classical system; in real time there were no written rules, this is just what was expected.[41]

Of course once we allow for the central bank to intervene there is always the possibility that they could do this in the "wrong" way – against the rules – by trying to prevent the domestic money supply from responding to gold flows. In modern parlance this is called *sterilization.* But this could only be conducted for short periods of time, because eventually a country would run out of gold. Moreover, it was inconsistent with the rules, and prior to 1914 countries treated the gold standard as an institution that could not be discarded.

Evaluation The advantage of the gold standard is that it ties the world price level to the world supply of gold. This is an advantage because it prevents inflation (unless there is a gold discovery). Notice that the gold standard does not prevent fluctuations in the price level, however. What it does produce is long-run price stability.

But the gold standard is also problematic. It ties the world money supply to the production of a commodity. There is no inherent reason why the growth in gold supplies will be related to the needs of international liquidity. When gold discoveries are rare, the world supply of gold will not increase as fast as real income. This would be deflationary. Between 1873 and 1896, the frequency of gold discoveries was rare while economic growth was rapid. With the world demand for money exceeding the growth in its supply, the price level had to fall. That

[41] Of course playing by these rules meant augmenting gold flows not sterilizing them. This is very unpopular, potentially, as it involves sacrificing internal balance for external balance.

explains why price levels fell dramatically (53% in the US). Gold rushes led to the opposite effect.

Bimetallism When gold production lagged the growth in output the pressure on the world price level was deflationary. This had two effects. One was to increase the reward to further gold discovery, and to conservation of non-monetary uses of gold. The second effect was to increase the use of silver as a medium of exchange. By maintaining a fixed rate of exchange between gold and silver (say 15.5 to 1 as in France in the first half of the 19th century), silver production could make up for shortages of gold. As gold became scarcer the return to silver discoveries increased, because countries on bimetallic standards exchanged gold and silver at the fixed rate. Two types of money circulated simultaneously.

Notice that as long as the production of gold and silver did not get too far out of line bimetallism was feasible. But if gold became scarcer a country like France could quickly run out of one metal. Agents would go to the mint in France with silver and exchange it for gold. Of course, the mint could raise the price of gold in terms of silver. But this would reduce the return to silver miners, and it would reduce the world supply of media of exchange. Political interests that supported debtors would agitate in favor of silver coinage.

In fact the saga of the *Wizard of Oz* is,[42] in fact, an allegory about bimetallism. Wholesale prices in 1890 were about 55% of their 1869 level. Real output had grown during the 1870's by about 5.0%, while the stock of money grew only by about 2.6%. The problem was the lack of specie. Congress responded, in 1890, with the Sherman Silver Act, which began coining some silver in limited amounts (at a rate far above 16 to 1). This caused panic that the US would leave the gold standard, and Grover Cleveland achieved its repeal. Unemployment in 1892-96 was over 12%, so when Democrats met in Chicago their cry was for bimetallism at 16 to 1.[43] Recall Bryan's call that "Thou shall not crucify mankind upon a cross of gold."

And thus Baum told the story in his book. Oz, in fact, stands for ounces of gold. Dorothy

[42] See Hugh Rockoff, *JPE*, 98, 4, August 1990.

[43] The market rate at that time was closer to 31 to 1. Hence free coinage would have led to some outflow of gold. But notice that free coinage really amounts to increasing the total stock of specie (like painting silver gold), which undoubtedly was in short supply at the time.

from Kansas is the honest American from the heartland, and the Scarecrow stand for farmers.[44] The Tinman is the worker and the Cowardly Lion is Bryan.[45] The Wicked Witch of the East is Wall Street – the advocates of tight money and most especially Grover Cleveland. The Wicket Witch of the West is drought – at that time ruining farms in Kansas and Nebraska (hence, destroyed by water).[46] Emerald City is Washington, where people must wear green shaded glasses; thus they are forced to see the world through the shade of money. Toto stands for "teetotaler," the prohibitionists, who agreed with the populists on silver. The Wizard is really just a man, and his solution for Dorothy's problem – the balloon – vanishes like hot air. Dorothy is eventually saved by Glinda, the Good witch from the South, the region the populists hoped to ally with, and all she has to do is click together her *silver* shoes. Thus when silver is coined along with gold all is solved.

While free coinage of silver would offset the shortage of gold, and would have offset the deflationary effects of the gold standard in the last quarter of the 19th century, the gold standard was on its way. This was partly due to the difficulties of operating a bimetallic standard.[47] Another factor was increased gold discoveries – primarily South Africa – that alleviated the gold shortage at the appropriate moment. It was also due to the fact that the strongest economy, Britain, had adopted the gold standard, and countries that traded with Britain found it advantageous to adopt the same system to facilitate trade.

Was Bryan foolish? Who could have known that gold would be discovered? And wasn't the money stock too tight? Moreover, given the size of the US is it unlikely that the market rate could have stayed at 31-1 while the US coined at 16-1. The price of silver would have risen and the world money stock as well.

[44] He complains of no brain – not understanding what the moneymen from the east tell him – but of course he finds that he has one by the end.

[45] This apparently is because by 1900, in his second race with McKinley, Bryan no longer fought the bimetallism issue. Baum is thus picturing him as a coward.

[46] Would inflation serve like water to end the problems of the farmer? What about the Fisher effect?

[47] Though these were probably overstated given the size of the US. The so-called instability of a bi-metallic standard was illusory. Two legs are better than one.

4.0.1. Interwar Period

Between the wars the gold exchange standard was developed. This basically differed from the classical gold standard in several ways.

- withdrawal of gold coins from circulation and concentration of gold stocks in central banks
- emergence of the dollar as a second reserve currency
- reduced wage and price flexibility especially in US and UK
- central banks no longer wished to play by the rules – emergence of popular democracy

Mundell argued that the primary failure of this system in the interwar period was too low a price of gold. The dollar price of gold was left unchanged even though prices had risen. A higher gold price would have increased liquidity.

Example 3 *Too low a price of gold*

What happens when the gold price is set too low relative to the price level, as in the US and UK after WW1? Too low a price of gold means that there is excess demand for gold domestically, and that the gold flow will be negative – our prices are too high so we lose gold to our trading partners. Notice that adjustment requires the gold stock, and hence domestic prices to decline. If this happens then the relative price of gold returns to its equilibrium level. The problem in the interwar period is that prices were rigid downward, so the deflationary pressure lead to unemployment

Consider figure 15. Initially the gold stock is at G_0. Given the price level equilibrium requires a price of gold high enough so that the relative price is $\left(\frac{P^G}{P}\right)^*$. But suppose that the gold price is set too low, so that we are at $\left(\frac{P^G}{P}\right)^0$. Then there is excess demand for gold at this low price, and we are running a balance of payments deficit – the flow of gold is negative, equal to the distance yx. To relieve the excess demand – to make the gold standard function

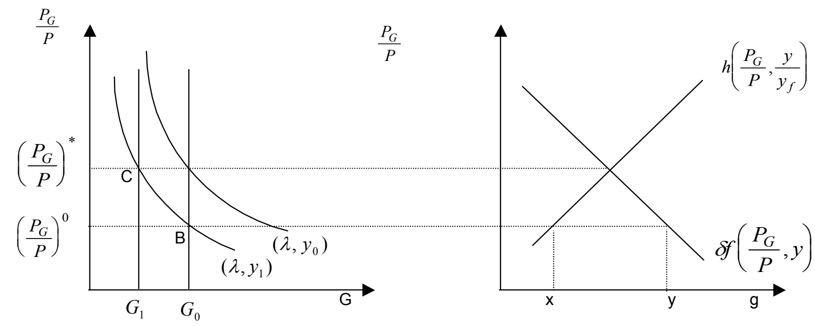

Figure 15: Too low a price for gold

at all – the demand for gold must fall, either by a decrease in income (as in the figure) or a decrease in λ.

- By the rules of the gold standard we must eventually get to point C. Here the price level has fallen so that we are once again in stock-flow equilibrium. This happens as $g < 0$ causes $G_0 \to G_1$. The fall in gold leads to lower prices and we end up at C. Deflation saves the day.

- "Playing by the rules" the CB could tighten monetary policy to hasten the fall in prices so that less gold must leave the country. That is, policy reinforces the deflation.

- But if the CB is unwilling or unable to sacrifice internal balance for external balance then problems arise. If they sterilize the gold flow, prices do not fall, and we stay at B, losing gold. This cannot go on forever. If they conduct expansionary policy to combat the recession it is even worse.

But contrary to this are two points: first, the gold cover ratio was little changed; second, foreign exchange reserves could substitute. But the problem of foreign exchange reserves was that they only work when gold prices are stable. If gold prices were to be changed, central banks may try to exchange foreign exchange for gold.

Remark 4 *This is the Triffin dilemma. Increased liquidity needs are met by foreign exchange reserves. This works as long as people are willing to trust the gold stocks. But this expansion threatens the credibility of the system.*

Two other problems are important. First, most gold was concentrated in three countries – US, France, and Germany. From 1927 to 1930 the share of world gold in these three countries rose from 56% to 63%. This pushed deflation on the rest of the world. Second, central bank credibility declined as they no longer played by the rules of the game. This threatened confidence in the system. Increasing pressure for domestic stabilization. Central banks started to liquidate foreign exchange reserves – the ratio of foreign exchange reserves to gold fell from 37% in 1930 to 11% by the end of 1932. More deflationary pressure.

Moreover, there were insufficient discoveries of gold. Perhaps the system was saved earlier by fortuitous gold discoveries. The gold rush, after all, was not caused by prospectors, but by the accidental discovery at Sutter's Mill. Periodic gold discoveries offset the deflationary bias in the system. By the interwar period luck had run out. Only pyramiding reserves could help, but this leads to Triffin dilemma.

Eventually, US left gold, then revalued gold to $35 per ounce, but the system really collapsed.

4.0.2. Bretton Woods

The Bretton Woods system was not exactly the gold standard, but what is called the gold exchange standard. All countries fixed their currencies in terms of the dollar, while the dollar was fixed to gold. Central Banks held reserves in the form of dollars, but these were claims on US gold supplies.

Four differences from the previous system.

- pegged exchange rates became adjustable subject to the existence of a fundamental disequilibrium
- controls on capital flows to add credibility given that monetary policy would be driven by domestic concerns
- IMF created
- limits imposed on private holdings of gold

These innovations dealt with the problems that were thought to have plagued the previous system – deflationary bias from overvalued pegs, a mechanism to contain destabilizing capital flows, a global liquidity shortage, and a mechanism to influence governments that were pursuing unstable policies.

But these innovations did not deal with the fundamental difficulty – how to cope with increased liquidity needs. Fast economic growth increased liquidity demands. Inelastic gold supply and limited price flexibility made this more acute. Only pyramiding foreign exchange could cope – now essentially increased dollar holdings.

The problem with this system again had to do with international liquidity. If the rest of the world is growing faster than the US – which was the result of the greater reconstruction after WW2 and postwar parities – then they would experience more rapid increased in the demand for money. Given their fixed exchange rates with the dollar, the only way for them to expand their money supplies was to increase their holdings of dollars. Hence, the only way for world liquidity to go up was for the US to run balance of payments deficits. This would increase the holdings of dollars in the rest of the world.

The world was enmeshed in the "Triffin dilemma." To keep the supply of international liquidity equal to demand the US had to run balance of payments deficits. But the increasing stock of dollars held outside the US increased the likelihood that Fort Knox had insufficient gold to back up the dollar. If all foreigners tried to cash in their dollars we would not be able to meet demand. If the US acted to curb deficits, on the other hand, there would be insufficient liquidity in the rest of the world.

Note the difficulty:

- If the US continued to run deficits it supplied liquidity but this threatened the gold backing
- If the US cut back on deficits there would be a liquidity shortage
- If the US raised the price of gold it would be going back on its commitment and threaten

the system.[48]

To meet this problem two possibilities emerged. First, an artificial asset could be created to act as a reserve currency, "paper gold." This was the function of SDR's issued by the IMF. Second, the US could devalue. In fact, both occurred.

The latter was the ultimate demise of the system. US inflation meant that other countries had to import inflation. Those that did not want this threatened to exchange their dollars for gold. But there was not enough. We ended up with the non-system that has existed since 1973.

[48] If a country does it once it may do it again. Hence, countries would prefer to hold gold rather than dollars. But then the system would revert to the gold standard from gold exchange.