

NBER WORKING PAPER SERIES

MEASURING MARKET INTEGRATION:
A MODEL OF ARBITRAGE WITH AN
ECONOMETRIC APPLICATION
TO THE GOLD STANDARD, 1879-1913

Gauri Prakash
Alan M. Taylor

Working Paper 6073

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 1997

We thank Lawrence Officer for helpful conversations and a brilliant book, and for guiding us to the *Financial Review*. We thank Eugene Canjels, Claire Waters, Mark Watson, Christopher Udry, Jan Tore Klovland, and Nathan Balke for help and useful comments. We also thank seminar participants at Rutgers University and conference participants at the conference on Integration of Commodity Markets in History held in Lund, Sweden, in March 1997. The usual disclaimer applies. A copy of the data used in this study is available from the authors upon request. This paper is part of NBER's research program in the Development of the American Economy. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

© 1997 by Gauri Prakash and Alan M. Taylor. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Measuring Market Integration: A Model
of Arbitrage with an Econometric Application
to the Gold Standard, 1879-1913
Gauri Prakash and Alan M. Taylor
NBER Working Paper No. 6073
June 1997
JEL Nos. N1, F3
Development of the American Economy

ABSTRACT

A major question in the literature on the classical gold standard concerns the efficiency of international arbitrage. Most authors have examined efficiency by looking at the spread of the gold points, gold-point violations, the flow of gold in profitable or unprofitable directions, or by tests of various asset market criteria, including speculative efficiency and interest arbitrage. These studies have suffered from many limitations, both methodological and empirical. We offer a new methodology for measuring market integration, based on a theoretical model of arbitrage applicable to any type of market. The model is econometrically tractable using the techniques of threshold autoregressions. We study the efficiency of the dollar-sterling gold standard in this framework, and we radically improve the empirical basis for investigation by compiling a new, high-frequency series of continuous daily data from 1879 to 1913. Using data at this frequency we can derive reasonable econometric estimates of the size of transaction-cost bands (as compared with direct cost estimates). We can also estimate the speed of adjustment through which disequilibria (gold-point violations) were corrected. The changes in these measures over time provides an insight into the evolution of market integration in the classical gold standard.

Gauri Prakash
760 Spinnaker Court
Secaucus, NJ 07094
gauri@merle.acns.nwu.edu

Alan M. Taylor
Department of Economics
Northwestern University
2003 Sheridan Road
Evanston, IL 60208-2600
and NBER
amu@nwu.edu

1 Introduction:

Gold Point Controversies

Our paper is concerned with assessing the degree of market integration in the dollar-sterling foreign exchange market in the late nineteenth and early twentieth centuries. The focus of our work, therefore, cannot be considered original—the study of the Anglo-American foreign exchange market has been one of the most actively pursued avenues of research in economic history. Several research papers from recent years illustrate the continuing attraction of this subject, and the key controversies. A major issue is whether the gold-standard regime was stable and efficient, and tests of this often center on whether the gold points bounded the market exchange rate. The gold points, the exchange rates at which gold arbitrage would be expected to commence, then need to be estimated to facilitate the test. Thus, the literature has frequently focused on getting this estimation right.

The conventional wisdom had long been that the rapid and *efficient* adjustment of the exchange rate under gold point arbitrage kept the dollar-sterling exchange stable (Officer 1996, 233): in this view, large deviations from par supposedly provoked gold flows sufficient to keep the rate always within the gold points (Cole 1929; Einzig 1970). However, influential revisionist contributions by Morgenstern (1959) and Moggridge (1972) sought to depict the classical gold standard as *inefficient*. Their estimates of gold points, based on transactions costs, suggested a narrow band, one frequently “violated” by the actual movement of the exchange rate, even in monthly time series. This approach was revisited by Clark (1984), with similar conclusions using refined estimates of transactions costs, confirming the frequent violations of the estimated gold points, and, even more disturbingly, persistent violations that spanned several monthly periods. Further consternation was caused by Clark’s finding that gold flows frequently did not correlate with arbitrage opportunities. Similar findings arose in an analysis of pure exchange-rate time-series behavior in a probabilistic model by Spiller and Wood (1988).

Standing up in defense of the conventional wisdom, Officer has produced a large body of research papers (1983; 1985; 1986; 1989; 1993), culminating

in a seminal book (1996). This corpus of work seeks to re-establish the remarkable efficiency of the dollar-sterling exchange over a large span of gold standard experience (1790–1931), and, in particular, during the heyday of the classical gold standard (1879–1913). Officer finds much at fault in the revisionist literature. For example, the major studies (Morgenstern; Moggridge; Clark; Spiller and Wood) all incorrectly use the cable transfer rates of exchange when the dominant arbitrage instrument was the demand bill (or sight bill). In addition, tests based on gold flow data were suspect given the known weaknesses of that data. Most importantly, all used *ad hoc* gold point estimates based on secondary sources. Officer’s response was to meticulously recalculate gold points from first principles, for the right instruments, and his summary volume lends considerable weight to the restoration of the conventional wisdom.

In a long run context, the conventional view also makes a good deal of intuitive sense, once the years of the classical gold standard are seen in historical perspective. It is clear from Officer’s quarterly data spanning almost one hundred and fifty years that, by the late nineteenth century, the dollar-sterling exchange had reached its peak level of stability. The exchange operated in the midst of an evolving gold standard. In technological terms, and perhaps in terms of learning by agents, by the turn of the century the market had reached its highest degree of sophistication: wide swings in the exchange rate were scarcely seen, as compared with experience before 1850, as shown in Table 1.

It should then come as no surprise that historians also place in this period, circa 1879 to 1913, the likely high-water mark of international capital mobility in the modern era. Indeed, by some measures, it is only in very recent years that measures of global financial market integration support the idea of a restoration of international capital-market integration comparable to historical experience under the classical gold standard (Eichengreen 1992; Obstfeld 1995; Obstfeld and Taylor 1997b; Zevin 1992).

In the face of an ongoing debate concerning the operation of the dollar-sterling exchange—and the classical gold standard more broadly—we believe that significant new contributions to the literature can be made in at least two ways. First, we can try to better formalize the theory of arbitrage in

Table 1: The Dollar-Sterling Standard: Summary Statistics, 1791-1966

Period	percent sterling premium over parity				
	mean	mean absolute	about mean	s.d. about zero	s.d. extremum
1791-1800	-2.70	4.55	5.06	5.75	-14.58
1801-10	3.46	4.17	3.48	4.93	9.52
1811-20	0.97	4.57	6.19	6.27	19.90
1821-30	1.23	2.01	2.06	2.40	-5.55
1831-40	-0.72	1.47	1.87	2.01	-6.10
1841-50	-0.73	1.11	1.26	1.46	-3.60
1851-60	0.42	0.65	0.68	0.80	-2.24
1861-70	0.32	0.87	1.20	1.25	-3.13
1871-80	-0.16	0.37	0.44	0.47	-1.09
1881-90	-0.19	0.33	0.36	0.41	-0.90
1891-1900	0.02	0.25	0.30	0.51	-0.61
1901-10	-0.03	0.14	0.19	0.19	-0.49
1911-14	-0.04	0.12	0.15	0.15	-0.28
1919-25	-0.12	0.24	0.27	0.29	-0.60
1925-31	-0.14	0.22	0.20	0.25	-0.43
1950-66	0.02	0.26	0.32	0.32	0.65

Sources: Officer (1996, 103).

the markets concerned; and, second, we might bring more suitable high-frequency data to bear on the question. In this paper, we make progress on both fronts, but with an approach that is radically different to any previous method. It is appropriate, then, to spell out the specific ways in which our analysis departs from the “state of the art” in the current literature.

2 Current Methodology: Estimating Gold Points

One key concept in the literature is the notion of the *gold points*, the size of exchange-rate deviations from mint parity which induce flows of gold to begin as private arbitrage becomes profitable. The other key concept is the volatility or variability of the exchange rate, or the average size of deviations from parity. Both approaches have been a staple in a very long literature dedicated to a better understanding of the functioning and development of the dollar-sterling exchange, including works by diverse financiers, economists, and historians such as Moggridge (1972), Davis and Hughes (1960), Bloomfield (1959), Morgenstern (1959), Cole (1929), Spalding (1915), Seyd (1868), and Goschen (1861). Officer’s (1996) book summarizes and vastly expands upon decades of research dedicated toward two related goals: the estimation and measurement of the gold points themselves, and the use of gold-point estimates to assess market efficiency. The latter enterprise uses the gold-point estimates to identify transactions costs, violations of the gold points, violations of interest parity, and the like.

One deficiency in all this work, identified by Officer (1996, 187) is a curious dichotomy which finds studies of the gold points and the movement of the exchange rate almost always divorced from each other. Aside from its use in identifying gold-point violations, the measurement of the gold points in itself is usually a static task of analysis, unrelated to the dynamic movements in the exchange. Conversely, studies of exchange-rate volatility almost never take into consideration the existence of gold points within the bounds of sample variation. Officer (1996) draws attention to this, and attacks the problem with studies of the exchange rate within and outside the gold point spread.

Our paper is certainly driven by the same ambition: to see the analysis of the gold points and the dynamics of the exchange rate placed in a unified formal framework.

A consequence of our analysis is that it yields a method of indirectly estimating the gold points, using implicit information on transaction costs embedded in the exchange-rate dynamics. We see this as an important benefit, and a useful cross-check on the reliability of both our model and the existing gold-point estimates based on more direct estimation. In fact, it is still an open question whether current gold-point estimates, based on direct estimates of transaction costs, are consistent with the actual historical behavior of arbitraging agents in the market. One consistency check, which has been used, is to see whether actual gold flows coincide with the implied gold-point violations derived from a particular time-series of cost estimates. This approach was suggested over one hundred years ago as a means to assess the workings of the exchange by Newcomb (1886, 281–282), but his suggestions were neglected at the time, and the most widely-used estimates of the gold points published in newspapers relied on less scientific, informal methods. The method has recently been applied by various authors (e.g., Clark 1984), but has run into practical difficulties. As noted by Officer (1996, 118), the bilateral gold flow data for many periods were, in Morgenstern's (1955) words, "worthless" for fine-scale analysis, and, in Goodhart's (1969) view, in need of corrections for four independent sources of error: shipping time, transit shipments via third countries, incomplete customs reporting, and clerical errors. Moreover, such corrections are feasible only after 1899, Officer argues. A major contribution of this paper is to show how to implement Newcomb's elegant test in a way which detects the point at which we see the onset of arbitrage behavior, but using only the time-series properties of the reliable exchange-rate (price) data, and avoiding the pitfalls of using the dubious gold-flow (quantity) data. Moreover, our method is in principle applicable to the arbitrage of any commodity, extending the range of the technique to the analysis of general problems in the study of market integration.

Our method is thus a complement to, rather than a substitute for, direct methods of cost measurement. Officer pursues gold-point estimation based on direct constructions of transactions costs from all their individual com-

ponents. He arrives at this method by discounting all the other available methods of calculation (Officer 1996, 117–121), which we now summarize. The alternatives considered are all the methods used elsewhere in the gold-point literature. Officer’s nine categories can be collapsed into four:

(a) Consult an Expert or Two, or Do Even More Averaging Across Time and Space. An “Expert” typically meant an experienced market participant such as a foreign-exchange dealer or a banker. This was a favorite method of contemporary studies of the gold points, e.g., by the British Parliament (1810, using testimony of merchants) or newspapers like the *New York Times* (1895, using dealers and houses) and *The Economist* (1925–31, using a bullion broker, Samuel Montagu and Co.). But such sources, even if consistent over time, may be unrepresentative, may not cover all cost components, and may not provide a full set of time series observations. Of course, a larger number of experts could be consulted, and a “consensus” estimate derived. This approach was often used by the *New York Times* and *The Economist*, but again yielded few usable observations. Most of the caveats just mentioned would still apply. Finally, one could average such estimates across time and space, as in Morgenstern (1959) and Moggridge (1972). This may have the benefit of averaging out errors, but one is then constrained to impose an unvarying gold point across time; and the different experts at different times might be using inconsistent methods of estimation.

(b) Find the Exchange Rate at Which Gold Flows. This is the notion behind the cross-check described above, due to Newcomb (1886). The method is elegant and has obvious theoretical appeal. It is based not on arduous and time-consuming cost calculations of every transaction component, but rather it is based on *actual behavior*. That is, gold points can be inferred using the principle of *revealed preference*. But the main problem is that the underlying quantity data (on flows of gold) are not as reliable as the price data (on the exchange rate), as we have already noted.

(c) Use the Exchange-Rate Maximum and Minimum. The advantage of this technique is the minimal data requirement. One only needs the exchange-rate series, which are data of acceptable quality. The main problem is the maintained assumption: that the exchange rate *never* exceeds the gold point, or, implicitly, that any gold-point violations are eradicated so “instantaneously” by arbitrage that we

never observe them. This is clearly a tenuous assumption: trans-Atlantic shipments of gold or sterling bills took seven to fourteen days in the period 1879–1913. There is no good reason why disequilibria could not persist for such spans of time, and thus be observed in the data—at least in *our* data, which is at a daily frequency. Of course, one could use weekly, monthly, or quarterly observations, but such averaging would, by construction, destroy data on the short-run dynamics of the exchange rate, and leave little information for assessing behavior far from parity when arbitrage corrections should occur over short horizons. Another problem is the use of long spans of data over which to take the maxima and minima: must we assume the gold points never change over time? Even a sophisticated econometric technique based on minima-maxima estimation with stochastic costs like Spiller and Wood (1988) is subject to this critique.

(d) Break Down All Individual Components of Transaction Costs and Estimate Directly for Every Time Period. Officer prefers this method. He discounts as “pure conjecture” efforts like that of Clark (1984) to take shortcuts, e.g., by projecting time-varying interest costs but using an invariant direct cost of shipping. He dismisses estimates of total arbitrage cost as unreliable, and insists that nothing can substitute for individual estimation of each cost component: interest, abrasion, insurance, freight, handling, mint charges, exchange rate commissions, and other miscellaneous charges. The benefit of this technique is that it makes no maintained theoretical assumptions and relies entirely on accurate measurement. One (obvious) problem with the technique is its intense data requirements. A less-obvious question is whether these direct cost estimates do indeed correspond to the cost-benefit parameters actually used by the arbitrageurs in question, an issue we soon confront.

It is a testament to Officer’s determination and great scholarship that he shows the feasibility of method (d) above. His arguments clearly refute the idea that loose inferences about the gold points in method (a) can be tolerated in the pursuit of any serious analysis of the dollar-sterling exchange. For example, the publication by *The Economist* of an invariant gold point spread (for several currencies, probably due to Ernest Seyd) in every issue from late 1877 to 1916 is rightly deemed unsatisfactory. The critique of shortcut techniques, like those that assume some components of costs, or use cost aggregations, are properly tested against the strict and unforgiving

ruler of repeated fine-scale measurement. Future generations of scholars will benefit from the detailed cost estimates Officer has constructed in his studies of the exchange.

However, even the best estimates still do not directly confront the question of how efficiently the exchange worked, and for that, as noted, we need to study the dynamics of the exchange rate as it relates to the gold points. Our model of exchange rate adjustment has this in mind, and, interestingly, it will display some of the advantages of the alternative methods (b) and (c) above, whilst trying to minimize their shortcomings (we hope). Like the gold-flow method (b), we will be using some maintained assumptions from theory, employing revealed preference to identify the gold points, following Newcomb (1886); but we will do this using only exchange-rate (price) data, without any need for the unreliable gold flow (quantity) data. So, like the maximum-minimum method (c), our model will require only price data, but, unlike that method, we will admit movements of the price both within and outside the gold-point spread. Lastly, method (c) has suffered from the use of monthly data. The maxima and minima of monthly averages may dramatically understate the movement of prices in a market where arbitrage times were calculated in terms of days for the Atlantic crossing. In contrast, we have constructed a new data set and we will use high frequency (daily) data that more closely correspond to the adjustment horizon in the actual market.

We arrive at our approach by the construction from first principles of an optimization model of gold arbitrage. The model relates gold flows to exchange rate movements. The basic ingredients are supply and demand equilibria for foreign exchange, and the solution of a profit-maximization problem for an arbitrageur. From this model we can derive the implied time-series process for prices. The price process then yields sufficient statistics to identify the convexity of arbitrageurs' cost functions, as measured by linear and quadratic terms. This allows us to infer the implicit transaction costs as revealed by the data.

We think that such a dynamic model properly bridges the gap between the theoretical working of the dollar-sterling exchange and standard measures of its efficiency such as the size and persistence of deviations, and the spread of the gold points. Moreover, such an arbitrage model is only one example

of how econometric analysis of price series can be used to assess market integration by revealing the implicit underlying costs facing arbitrageurs. For example, Prakash (1996) applies the method to rice markets in India, using weekly wholesale prices from the 1960s to the 1980s to estimate “rice points”; and Obstfeld and Taylor (1997a) present an application to purchasing power parity and the law of one price, in the spirit of Eli Heckscher’s notion of “commodity points,” using monthly panel data for quasi-tradable CPI components in 32 world locations from 1980 to 1995. Our method therefore invites application to the study of other price processes, and the comparison of implicit and direct costs estimates as part of the study of market integration.

3 A Model of Gold Point Arbitrage

In this section we develop a simple model of gold arbitrage applicable to the dollar-sterling exchange during the classical gold standard from 1879 to 1913. The model consists of two countries, a market for foreign exchange, and gold arbitrageurs acting between the two countries. We here present the theory, and in the next section show how to apply the model econometrically.

First we need to specify the foreign exchange market, and this requires consideration of the appropriate instrument used in the market and in arbitrage. It is now accepted that the dominant form of arbitrage under the classical gold standard was via the demand bill (or sight bill) denominated in sterling and drawn on London. Only later, in the interwar period, did cable or telegraphic transfers dominate the market. In earlier periods, the 60-day bill introduced in colonial times was the preferred instrument, except for a brief period in the mid-nineteenth century when a 3-day bill was used (Officer 1996, 113–115). The use of bills may seem surprising given the deployment of the first trans-Atlantic cable in 1866, but it appears to follow from high transaction costs involved in cable transfers and in securing forward cover for the duration of any gold shipment.¹

¹Recall that a demand (sight) bill in New York was a composite instrument of sorts, since it could be redeemed only in London (i.e., forward) but at a certain price. Thus forward cover was included in the bill, but it still was subject to discount vis-à-vis cable

We will now consider operations in the New York demand bill market. It is this market whose daily demand bill quotations we have collected from 1879 to 1913, so we attempt to formalize the equilibrium and dynamics of adjustment in this market.

Let the stock of bills in the “market” be B_t .² The price of these bills in U.S. dollars is simply the exchange rate on demand (sight) bills expressed in U.S. dollars per pound sterling, E_t . The first building block of our model is an expression of the market for such bills, written as a demand curve,

$$E_t = \xi - \eta B_t + u_t, \quad (1)$$

where ξ and $\eta > 0$ are demand parameters, and u_t is a shock to the demand curve at date t . This equation states that an increase in the quantity of bills B_t in the New York market leads to a fall in the price of bills E_t .³

rates due to the interest cost of shipping the bill to London. Thus, cable rates might have been more advantageous than demand bill rates in this period, but the thin market and high costs seem to have allowed the demand bill to maintain its dominance (Davis and Hughes 1960; Perkins 1975; Officer 1996, 60–63, 115). For this reason, some authors pedantically or jokingly refer to the classical gold standard as “the demand bill on London standard.”

²By “market” we mean New York, strictly, but the extent of this market could, of course, include the entire United States, or as much as might be linked to the New York market for sterling bills. Indeed, the development of the internal U.S. market and its relationship to the external market has also been a subject of considerable debate (Davis and Hughes 1960; Officer 1996).

³In what follows we will close the model with a supply relationship where changes in the stock of bills result from gold arbitrage. Technically, this was not the only source of bill supply. In principle, bankers in New York, for example, could create new bills for redemption in London *without* a corresponding gold movement to cover the bills. Instead, the banks would use the sale proceeds to acquire U.S. dollar assets in New York, and would cover the bills in London via the sale of British sterling assets. To model this process would require, however, a model of banks’ international portfolio choice, and to estimate it would require comprehensive data on bank portfolios. We have no such model and no such data. We think our simplifying assumptions reasonable, however, in certain respects. Our demand curve must be the demand curve that obtains in the long-run equilibrium as banks cannot without limit convert a portfolio from U.S. assets to British assets or vice versa. Second, if there is an optimal portfolio choice then deviations from that choice might properly be modeled as an equilibrium error, and we might suspect that the changes in the equilibrium portfolio occur at much lower frequencies than the daily actions in the gold arbitrage market. As such, our model still applies. Finally, even if the demand curve should instead be seen as a reduced-form price equation, embedding both

The time series process u_t is of concern, and it will be important in what follows. We have no simple priors on this process, except to say that in the long run it is probably not stationary. It may even have a deterministic trend, related to long-run trends in, say, the international trade in New York and its dependence on sterling bills, or possibly other structural factors relating to technical change or growth in the financial sector. Further, at daily frequency, u_t might reasonably be expected to be serially correlated.

The next building block of our model concerns the relationship of gold arbitrage to the stock of bills. Whenever arbitrage via demand bills takes place, the arbitrageur effectively swaps a demand bill in New York for gold in London, or vice versa, through shipments across the Atlantic. As Table 2 shows, revenues could be derived in this trade when the exchange rate E_t (the market price of demand bills in New York) diverged from its par value E^{par} , the latter given by the ratio of the fixed mint prices of gold in the United States (P^g , in dollars per ounce) and Britain (P^{g*} , in pounds per ounce), that is, $E^{par} = P^g / P^{g*}$. Provided such revenues exceeded transaction costs, the trade would be profitable.⁴

Table 2 relates the changes in the quantity of gold and bills in the typical transactions based on the account in Officer (1996, 111–113). Two points should be made. First, we assume that the New York arbitrageur takes profits in dollars in New York, so that in the case of gold export the revenue is taken out before gold is acquired for shipment.⁵ Second, note that in the case of export there is no interest cost, since profits are immediately realized,

supply- and demand-curve elements, it is still the case that our reasoning will remain valid so long as the parameter η has the correct sign, which is so long as the demand response is elastic relative to the supply response of banks creating bills. This is reasonable too, for if η had the wrong sign we would be in the neighborhood of an unstable exchange-rate equilibrium, given the arbitrage process we next describe.

⁴To follow our earlier footnote remark, banks which create bills but do not ship gold to cover the bills, engage in only the two bill trades shown in each half of Table 2. The gold trades do not occur. Instead, to complete the transaction they must buy and sell assets in each market. But as noted earlier, this kind of activity has long-run limits, and is constrained by banks' international portfolio targets. We consider it a second-order element of gold-standard arbitrage purely construed, and we omit it in the present analysis. At the margin we are focusing on the arbitrage processes described in Table 2.

⁵That is, the arbitrageur exports only enough gold to cover the bills that need to be redeemed in London, and keeps the remaining dollars as net revenue in period $t = 0$.

Table 2: Gold Import and Export Via Demand Bill.

Export of one ounce of gold, with $E > E^{par}$

time	place	action by arbitrageur	gain	loss
$t = 0$	New York	sell demand bills	\$ EP^{g*}	bills $\pounds P^{g*}$
$t = 0$	New York	buy gold	gold oz. 1	\$ $E^{par} P^{g*}$
$t = T$	London	sell gold	$\pounds P^{g*}$	gold oz. 1
$t = T$	London	redeem bill	bills $\pounds P^{g*}$	$\pounds P^{g*}$
		Marginal Revenue	\$ $(E - E^{par})P^{g*}$	
		Bills Inflow	$\pounds +P^{g*}$	
		Gold Inflow	oz. -1	
		Interest cost in time	zero	

Import of one ounce of gold, with $E < E^{par}$

time	place	action by arbitrageur	gain	loss
$t = 0$	New York	buy demand bills	bills $\pounds P^{g*}$	\$ EP^{g*}
$t = T$	London	redeem bill	$\pounds P^{g*}$	bills $\pounds P^{g*}$
$t = T$	London	buy gold	gold oz. 1	$\pounds P^{g*}$
$t = 2T$	New York	sell gold	\$ $E^{par} P^{g*}$	gold oz. 1
		Marginal Revenue	\$ $(E^{par} - E)P^{g*}$	
		Bills Inflow	$\pounds -P^{g*}$	
		Gold Inflow	oz. +1	
		Interest cost in time	$2T$	

Note: T is time for a one-way trans-Atlantic voyage. In this table non-interest costs are not shown, such as mint charges, assaying, freight, insurance. See text and the discussion of the model.

Source: See Officer (1996, 111–113).

but in the case of gold import, the time interest cost is that of a two-way Atlantic voyage, where a one-way trip takes a time T .⁶ This introduces the possibility of an asymmetry in costs, and hence in the gold points, due to different interest considerations in each direction.

The table shows that, in both directions, a transaction financed by the sale (purchase) of a one pound sterling demand bill is used to finance the export (import) of $E^{par}/P^g = 1/P^{g*}$ ounces of gold. Thus, the relationship between gold movements and changes in the stock of demand bills is given by

$$\Delta B_t = -P^{g*} \Delta G_t, \quad (2)$$

where ΔG_t is the gold *inflow* into New York, and, by definition, $\Delta X_t = X_t - X_{t-1}$ for any variable X_t .

The final building block of our model is the cost function for the arbitrage operation, which, together with the above revenue calculation, will determine optimal arbitrage. We will consider the cost function to be a convex (for simplicity, quadratic) function of the size of flows. As just shown, the quantities of bills and gold involved in arbitrage are proportional, so either might be used to measure the flow. We use the flow of gold ΔG_t as our measure. We will consider a representative arbitrageur. We suppose that this arbitrageur faces a cost function for transactions that depends on the quantity of transactions (the amount of arbitrage). In reality, with many arbitrageurs, this amounts to a definition of the supply curve of arbitrage services. Since arbitrage operates in both directions, the cost function will be a function of the absolute size of the flow $|\Delta G_t|$. For simplicity, let us suppose that the cost of a total gold flow of absolute size $|\Delta G_t|$ ounces at date t is given by the total cost, in U.S. dollars,

$$TC = a + b|\Delta G_t| + \frac{1}{2}c|\Delta G_t|^2. \quad (3)$$

This is a general technology where there are fixed costs a , an initial marginal cost b , and an increasing marginal cost at a rate c . In principle, the potential asymmetry in interest costs can be accounted for by varying the coefficients

⁶With gold import, the bills have to go to sale in London at time T , and the gold has to come back and be sold for the dollar profit which is not realized until time $2T$.

of this transaction technology in each direction. Other possible sources of cost asymmetry are differences in abrasion costs, insurance premia, freight rates, assaying charges, bank fees, or other cost components in each direction.

We now consider the arbitrage operation as a profit-maximization problem. Arbitrageurs will only move gold up to the point where marginal revenue equals marginal cost. Consider the case of gold import at time $t - 1$, with $E^{par} > E_{t-1}$. Suppose that $z > 0$ units of gold are already being moved. Consider the movement of an incremental ounce of gold. Based on Table 2 we know that the marginal revenue of a one ounce gold movement via demand bill is given by $MR = (E^{par} - E_{t-1})Pg^*$; and we know the marginal cost is $MC = b + cz$ from (3).

By equating MC and MR , and by a similar calculation for gold export, we obtain the optimal arbitrage flows $z = G_t - G_{t-1} = \Delta G_t$ as a function of exchange rate deviations $E_t - E^{par}$,

$$\Delta G_t = \begin{cases} -\frac{1}{c}((E_{t-1} - E^{par})Pg^* - b) & \text{when } E_{t-1} - E^{par} > b/Pg^*; \\ 0 & \text{when } |E_{t-1} - E^{par}| \leq b/Pg^*; \\ +\frac{1}{c}((E^{par} - E_{t-1})Pg^* - b) & \text{when } E^{par} - E_{t-1} > b/Pg^*. \end{cases} \quad (4)$$

Now, from (1) and (2) we know that gold flows and the exchange rate are related via

$$\Delta E_t = \eta P^{g^*} \Delta G_t + v_t, \quad (5)$$

where $v_t = \Delta u_t$. We think the process v_t , as the difference of u_t , likely to be a stationary process. In what follows, we will assume as a benchmark that v_t is white noise $N(0, \sigma^2)$. However, our arguments will generalize to other processes for v_t .

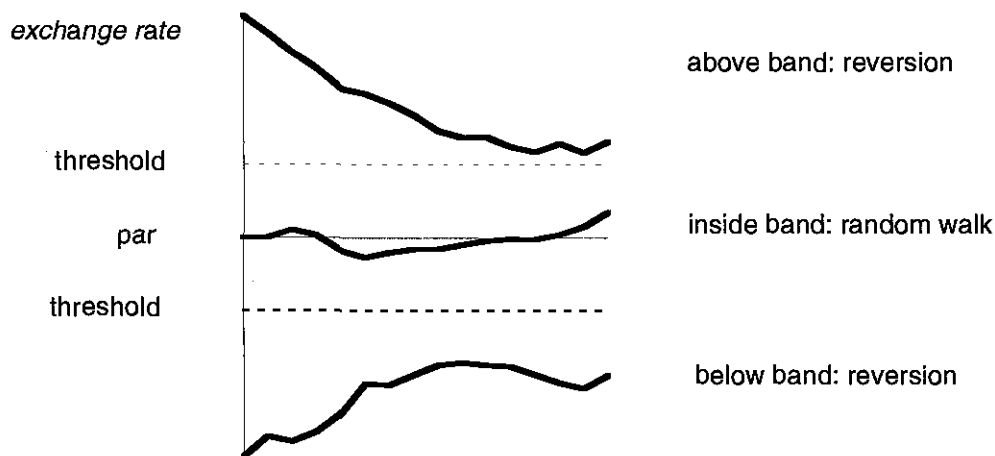
A solution follows from (4) and (5). We find that the dynamics of E_t are given by

$$\Delta E_t = \begin{cases} \lambda(E_{t-1} - (E^{par} + \gamma)) + v_t & \text{when } E_{t-1} > E^{par} + \gamma; \\ v_t & \text{when } |E_{t-1} - E^{par}| \leq \gamma; \\ \lambda(E_{t-1} - (E^{par} - \gamma)) + v_t & \text{when } E_{t-1} < E^{par} - \gamma; \end{cases} \quad (6)$$

where

$$\begin{aligned} \lambda &= -\eta(P^{g^*})^2/c; \\ \gamma &= b/P^{g^*}. \end{aligned} \quad (7)$$

Figure 1: Exchange Rate Adjustment in the Model



This model incorporates the simplest formulation of gold point dynamics. If the current exchange rate is outside the band given by the gold points $[E^{par} - \gamma, E^{par} + \gamma]$, then the exchange rate reverts toward the edge of the band (the nearest gold point) at a speed λ . Within the band, between the gold points, there is no reversion and the process follows a driftless random walk. These dynamics are depicted in Figure 1. We have also shown how the gold-point spread γ and the adjustment speed λ are related to the shape of the arbitrage cost function. An increase in the linear cost parameter b causes an increase in the gold point spread γ ; an increase in the quadratic cost parameter c causes a decrease in the speed of convergence λ .

The expression for λ shows that adjustment is instantaneous with an infinite exchange rate movement when $c = 0$, a case where marginal costs are constant. This is intuitively obvious: if marginal costs never rise, then an infinite amount will be shipped and the exchange rate will get back inside the gold points immediately.⁷ This case seems empirically less relevant, as such shipments are not seen, nor can costs be assumed to be linear forever.⁸

⁷For this thought experiment we need to imagine the period length shrinking to zero too, and perhaps per period costs therefore rising.

⁸However, such an adjustment dynamic may be the technical basis for the Spiller and Wood (1988) model where the exchange rate never goes beyond the gold point, as in

Otherwise, we are in a regime with a finite positive λ (the case $c > 0$) and adjustment takes place with a lag. This would seem empirically relevant given the known lags in the actual arbitrage process. The term γ shows that the gold points may be thought of as a pair of thresholds, delineating two regimes of exchange rate behavior. Inside the band, the exchange rate is unaffected by any reversion tendency since arbitrage is never profitable and remains inoperative. Outside the band, the exchange rate shows a tendency to revert to the band edge. The threshold is proportional to the minimal marginal cost of arbitrage operations (b), as expected.

This completes the model description. The next section shows that a simple econometric technique exists to estimate a model of the form (6), allowing us to identify the gold points and adjustment speeds of the arbitrage process using a method of revealed preference based solely on price dynamics.

4 An Econometric Methodology: Searching for Gold Points

To simplify the model, let us normalize by defining $x_t = E_t - E^{par}$, so that x_t is the deviation of the exchange rate from par. Then the model (6) is just a standard threshold autoregression (TAR) model,

$$\Delta x_t = \begin{cases} \kappa^{out} + \lambda^{out}(x_{t-1} - \gamma) + e_t^{out} & \text{when } x_{t-1} > \gamma; \\ \kappa^{in} + \lambda^{in}x_{t-1} + e_t^{in} & \text{when } \gamma \geq x_{t-1} \geq -\gamma; \\ \kappa^{out} + \lambda^{out}(x_{t-1} + \gamma) + e_t^{out} & \text{when } -\gamma > x_{t-1}. \end{cases} \quad (8)$$

For our purposes, we will assume e_t^{out} is $N(0, \sigma^{out2})$, e_t^{in} is $N(0, \sigma^{in2})$, $\lambda^{in} = 0$, and λ^{out} is the convergence speed outside the gold points. Furthermore, we will assume that convergence is indeed directed toward the thresholds and that there is no drift inside the band, so that $\kappa^{out} = \kappa^{in} = 0$. In this model, equilibrium obtains whenever x_t is within the gold points, so we still speak of convergence to equilibrium. Only now, equilibrium for x_t is anywhere in a band or interval $[-\gamma, +\gamma]$, and not just at the point 0. Within that band there is no “error correction” force at work, because there is no error

Method (c).

to correct. Note that a looser specification would leave λ^{in} unrestricted. We think it acceptable to posit a random walk for prices within the band as argued above. However, we allow nonzero constant terms κ^{out} and κ^{in} , because sample intercepts (means) will never be exactly zero.⁹

Our model is one of a large family of TAR($p;k,d$) models, which may be characterized by an arbitrary autoregressive length p , an arbitrary number of thresholds k , and an arbitrary delay parameter d (which denotes the order of lag used to determine within which threshold regime the current observation falls). The simple model we have proposed is a TAR(1;2,1) with some restrictions on the coefficients for symmetry about $x = 0$. Balke and Fomby (1997) call our model a “Band-TAR” model, for obvious reasons. There exists a simple parametric method of estimating λ and γ . Following Fanizza (1990), Balke and Fomby (1997), and Prakash (1996), we use a best-fit grid-search on the threshold parameter γ . Such a search requires an objective function, and some choices are available. Balke and Fomby (1997) search so as to minimize the residual sum of squares; Fanizza (1990) and Prakash (1996) maximize a likelihood function. We follow the latter approach here.¹⁰ Let $L_n(\lambda, \sigma)$ be the estimated log likelihood function of the reference (null) AR(1) model, and let

⁹We could also introduce asymmetry in the threshold variable, but this would greatly complicate the model: ours is a search algorithm on thresholds, so a grid search on N points of a line would become a search over N^2 points on a grid of upper and lower thresholds. Nonetheless, this would be technically feasible. However, in practice, current “mechanical” estimates of the gold points based on costs do not produce a consistent picture as regards the differences between the gaps between the two gold points and the par value. And even to the extent that there are gold point asymmetries in the literature, these are usually, small, less than 0.1% of par (Officer 1996, 174). For technical reasons, our grid search for the thresholds will be limited by the rounding in the data, as discussed in the next section. Most observations round to half a cent, which is just over 0.1% of par. Thus we doubt we have sufficient information to identify any asymmetries. Finally, we also note another econometric complication: to identify both upper and lower thresholds independently would require observations inside the band, and in *each* of the lower and upper bands. But for some years, even consecutive years, the exchange rate may remain close to par, and almost always on one side of par. Such data would be of no use in trying to identify the threshold on the other side of par. Thus, for several reasons we can defend our reluctance to estimate asymmetric thresholds. As a future exercise, or with other price series in other problems, such asymmetries could be estimated using our technique. But as a first approach here we retain symmetry of the thresholds relative to par.

¹⁰The methods are identical when we impose a simple OLS regression structure, as here, with no serial correlation or heteroskedasticity.

$L_a(\lambda, \sigma^{out}, \sigma^{in}; \gamma)$ be the estimated log likelihood function of the (alternative) TAR model (8) for a given γ . Estimation proceeds via a grid search on γ which maximizes the log likelihood ratio $LLR = 2(L_a - L_n)$. This is computationally simple, since for any given γ , maximum likelihood (ML) estimation of the TAR model amounts to an OLS estimation on partitioned samples, namely, sets of observations with x_{t-1} wholly inside or wholly outside the thresholds.¹¹

In order to test whether this kind of nonlinear specification is justified, we can use Tsay's (1989) F -test approach for testing general $TAR(p; k, d)$ alternatives against $AR(p)$ nulls. We will also employ Monte Carlo simulations to test our specific $TAR(1; 2, 1)$ against the $AR(1)$ alternative, since the $AR(1)$ may be viewed as an alternative model of adjustment, as per the literature on the gold standard as a target zone (see, e.g. Hallwood, Macdonald, and Marsh 1996). The likelihood ratio provides a basis for the Monte Carlo tests. This is a test of the specific two-threshold commodity-point TAR against the $AR(1)$ null, as compared with Tsay's general nonparametric test of $AR(1)$ versus alternatives with an indeterminate number of thresholds. However, a price is paid: care is needed since the parameter γ is not identified under the $AR(1)$ null and L is discontinuous. Thus standard inference is invalid, and we must proceed using a Monte Carlo method, as in Fanizza (1990).¹² Empirical estimates of the distribution of LLR under the $AR(1)$ null can be simulated, and used to test whether the null should be rejected in favor of our specific TAR alternative. The Tsay test is fully described in Appendix A and the TAR estimation and Monte Carlo method in Appendix B.

5 Data

Before we apply the model, it remains to describe the new data set we have compiled which will allow us to examine the workings of the classical gold standard at a much a higher frequency scale than any earlier work. Previous

¹¹For practical purposes, choices of γ with few observations in either partition are rejected (we chose 30 observations as a minimum).

¹²As shown by Balke and Fomby (1997), their "sup-Wald" statistic may be used to test for nonlinearity also, but again using Monte Carlo methods for the same reasons.

researchers have relied on monthly, quarterly, or annual data series. Even so assiduous a compiler of data as Officer (1996, chapter 6) could only garner quarterly data for the last two centuries as a whole, with monthly data for the years 1890–1906, 1925–31, and 1950–66. Although some weekly data are reported in certain sources (e.g., some documents of the National Monetary Commission), we decided to collect data at the highest feasible frequency for the entire period of the classical gold standard, namely daily data for the period 1879–1913. This covers almost the entire period of the dollar-sterling exchange under the classical gold standard, which began with the restoration of U.S. convertibility on January 1, 1879, and ended with the termination of U.K. convertibility in August 1914 (Officer 1996, 16, 43).

We went to the best primary source for this data, the *Financial Review*, which tabulated daily data on the New York exchange each year in an annual summary.¹³ Sight bill quotations are usually in a range, e.g. “4-85 86-1/2” meaning between 4.85 and 4.865 on that day. We convert such ranges to a midpoint. There is considerable rounding in the quotes: often the finest gradation is one half cent.¹⁴ This minimizes the information in the data, and effectively shrinks the meaningful range of points in the search algorithm to find the best-fit threshold (a blessing, at least, in terms of computational cost). The data covers every trading day, that is, every day except Sundays, holidays, and a few exceptional days on which the exchange was closed. We discard non-trading days and perform time series analysis on the series of price quotes for trading days only. Thus, we have about 300 daily observations on the New York price of sterling demand bills in each year from

¹³We thank Lawrence Officer for suggesting this source to us. The tabulations covered sight and sixty-day bills; we compiled both series, but only the information on sight bills is used here. Note that these are “posted rates” of banks, taken from the weekly data published by the same company in their *Commercial and Financial Chronicle*. The weekly publication also published “actual rates” for transactions, but these were not summarized for the annual review. We do not know which rates were more relevant for arbitrageurs. However, based on inspections for selected years it is hard to see any systematic tendency for “posted rates” to diverge from “actual rates,” except for a small difference in levels. We thank Jan Tore Klovland for pointing out this difference. In principle, one could comb the weekly publication for the “actual rate” daily observations, but this would require handling fifty-two times as many publications, as this proved beyond our scope. Instead, we took the thirty five annual summaries and entered the daily data on “posted rates.”

¹⁴Specifically, before 1904 the finest distinction is 0.5 cents, thereafter 0.05 cents.

1879 to 1913. This represents an enormous amount of data for time-series analysis: over 10,000 observations over the entire sample. With such a data set, we should have adequate information to identify parameters of interest and how they change over time.

6 Results

The results of applying the maximum likelihood estimation method to our data are shown in Table 3. We employ various sample periods: sequential disjoint decades, quinquennia, and triennia. We also tried using annual data, but the small number of observations proved to be a problem—not for accurately estimating OLS parameters, but because in certain years the exchange rate remained in a tight interval almost wholly inside the gold points. Such samples cannot provide enough information for the estimation of gold points, so we moved to longer sample periods. The table displays the estimated gold point threshold γ (in U.S. dollars) given by a threshold search over quarter cent values (\$0.0025) from zero to ten cents. Also shown are the number of observations T , the estimated speed of convergence λ outside the band and its t -statistic, and the estimates of σ^{in} and σ^{out} .

Our results provide a very different view of the dollar-sterling exchange from the perspective of the revealed preference of the actors, rather than from the viewpoint of measured transaction costs. It appears that the thresholds or implicit gold points varied a great deal over time, from about four cents early on in the 1880s to less than one cent around 1910. The speed of convergence also varies a great deal, but with no clear pattern. It is usually in the range of 2% to 5%, suggesting a half life of deviations between two and four weeks. There is, however, considerable variation in both parameters, though this is most noticeable in the triennial data when small samples might still be a problem. Nonetheless, some of the variation correlates with well known historical phenomena. For example, the threshold value is initially high, then small in the late 1880s, but then it levels off in 1890–92, a period of worldwide financial crisis. The threshold is even higher in 1893–95, when the U.S. silver controversy also troubled the exchanges. And then, in 1908–10, the threshold increases, just after the great 1907 U.S. financial crisis. Such

Table 3: Maximum Likelihood Estimation of TAR Model

years	T	gamma	lambda out	t out	sigma in	sigma out
(a) full sample						
1879-1913	10626	0.0375	-0.266	4.7	0.00200	0.00274
(b) decades						
1879-1888	3050	0.0150	-0.021	4.7	0.00279	0.00229
1889-1898	3034	0.0225	-0.019	2.3	0.00220	0.00132
1899-1908	3028	0.0200	-0.080	3.1	0.00171	0.00305
1909-1913	1514	0.0075	-0.038	3.4	0.00111	0.00124
(c) quinquennia						
1879-1883	1523	0.0400	-0.739	4.9	0.00259	0.00302
1884-1888	1527	0.0100	-0.019	3.4	0.00277	0.00207
1889-1893	1516	0.0175	-0.033	3.4	0.00299	0.00174
1894-1898	1518	0.0250	-0.012	1.1	0.00155	0.00113
1899-1903	1510	0.0150	-0.030	2.7	0.00154	0.00111
1904-1908	1518	0.0125	-0.095	3.4	0.00166	0.00384
1909-1913	1514	0.0075	-0.038	3.4	0.00111	0.00124
(d) triennia						
1878-1880	610	0.0175	-0.025	2.4	0.00312	0.00228
1881-1883	913	0.0400	-1.532	6.0	0.00269	0.00252
1884-1886	917	0.0300	-0.055	0.9	0.00286	0.00189
1887-1889	913	0.0175	-0.012	1.2	0.00194	0.00122
1890-1892	911	0.0175	-0.041	3.1	0.00289	0.00163
1893-1895	912	0.0250	-0.011	0.9	0.00264	0.00135
1896-1898	908	0.0025	-0.008	2.1	0.00112	0.00154
1899-1901	906	0.0150	-0.030	1.9	0.00172	0.00119
1902-1904	908	0.0150	-0.055	3.2	0.00133	0.00104
1905-1907	910	0.0125	-0.105	3.2	0.00195	0.00397
1908-1910	909	0.0150	-0.997	3.3	0.00128	0.00190
1911-1913	909	0.0075	-0.058	3.0	0.00098	0.00136

Notes: See text.

historical phenomena could be expected to raise questions as to the credibility of the gold standard, perhaps discouraging arbitrage via reputation effects. More directly, arbitrage thresholds could be shifted via the use of *gold devices*, the tools used by central banks to manipulate the gold points in times of crisis (see Bloomfield 1959).

It is tempting to try to assign historical events to these periodic movements in the estimated parameters, but we would not choose to stress this approach too much. Rather, we wish to point out that evidence contained purely in the price process suggests a foreign exchange market evolving across time, with arbitrage subject to time-varying costs. This does not imply an unstable or an inefficient gold standard, *pace* the revisionist critique. We think this evolution consistent with rational profit-maximizing arbitrage, as in the core of our model, and still supportive of a stabilizing effect of arbitrage given time and cost constraints on the speed of adjustment. By definition, our model allows gold point “violations” of a sort, but these disappear as disequilibria are dissipated subject to transaction costs.

Finally, we should note one other feature of the estimation process. The residual variance estimates σ^{in} and σ^{out} both decline across time. Since they are interpreted as representing shocks to the demand for bills, this may be seen as tentative evidence of an evolution towards a more stable exchange resulting both from a more stable external environment (external to the exchange, that is), in addition to a more stable exchange resulting from tighter gold points or faster adjustment speeds. Both forces can be discerned as being at work in reducing the overall dispersion of the exchange rate over the years as noted by Officer (Table 1).

Is our model justified? We have proposed some specification tests, and Table 4 contains some diagnostic information on these tests of the model. The left hand columns show that Tsay’s F -test for nonlinearity overwhelmingly favors the TAR model against a simple AR(1) alternative. The column under $p1$ shows the Monte Carlo tests of our specific TAR model against an AR(1) alternative.

We think these tests allow us to consider our TAR model a satisfactory description of the exchange-rate process under gold arbitrage. The first set of Monte Carlo tests do show that the TAR specification is superior to a pure

Table 4: Specification Tests of TAR Model

	TAR LLR	--- Tsay --- F	---- p	Monte Carlo p1	p2
(a) full sample					
1879-1913	46.2	58812.2	0.00	0.16	-
(b) decades					
1879-1888	65.6	23617.7	0.00	0.04	1.00
1889-1898	314.2	27196.1	0.00	0.04	1.00
1899-1908	277.7	14241.1	0.00	0.11	0.00
1909-1913	15.2	6687.3	0.00	0.12	0.71
(c) quinquennia					
1879-1883	30.4	10869.5	0.00	0.01	1.00
1884-1888	63.5	11173.4	0.00	0.03	0.72
1889-1893	218.1	12376.2	0.00	0.04	1.00
1894-1898	66.1	24863.6	0.00	0.01	0.00
1899-1903	76.0	10912.3	0.00	0.04	0.03
1904-1908	457.7	29.9	0.00	0.19	0.20
1909-1913	15.2	6687.3	0.00	0.11	0.71
(d) triennia					
1878-1880	31.0	3803.5	0.00	0.01	1.00
1881-1883	28.5	3414.4	0.00	0.01	0.90
1884-1886	41.2	6503.2	0.00	0.02	1.00
1887-1889	98.2	14250.2	0.00	0.01	0.69
1890-1892	144.1	7808.0	0.00	0.02	0.66
1893-1895	192.5	6457.6	0.00	0.01	0.14
1896-1898	17.9	10228.5	0.00	0.03	0.00
1899-1901	58.3	6705.5	0.00	0.02	0.60
1902-1904	25.0	7704.2	0.00	0.03	1.00
1905-1907	226.1	34.1	0.00	0.12	0.05
1908-1910	26.7	2720.0	0.00	0.16	0.05
1911-1913	54.0	1.2	0.30	0.10	0.84

Notes: Monte Carlo results from $n = 1,000$ draws. $p1$ is versus the AR1 null, and $p2$ is versus the TAR null with a constant threshold equal to the estimated threshold from specification (a). See text.

AR specification in most periods. However, another question still remains: what kind of TAR model is appropriate? In particular, our choice of sample periods suggests a time-varying threshold over the period 1879 to 1913. But is such a structural instability present? According to the traditional viewpoint there was a virtually static gold point over the entire classical gold standard period. Although no one takes seriously claims such as *The Economist's* that this gold point was literally invariant, most authors do, nonetheless find very little change in their estimates of gold points from year to year, with perhaps only the occasional small change to reflect changes in shipping or interest charges. Given our remarkable high-frequency data, we are in a good position to test the traditional hypothesis, and we proceed again by simulation. We assume the exchange rate follows the process given by the TAR fitted to the full-sample data from 1879 to 1913, and we generate simulated data from this specification, as in panel (a) of Table 3. For each such draw, we then estimate the time varying TARs shown in panels (b), (c), and (d), and we use Monte Carlo methods based on the *LLR* statistic to see whether such alternatives are preferred to the null of a time-invariant TAR. These results are shown in the column labeled *p2* in Table 4. The results show that it is impossible to maintain the assumption that the threshold of the process was stable across all periods, as in many cases the period-specific TAR produces a much better fit at conventional significance levels. Thus, our results also suggest a new view of the the classical gold standard, where the parameters of the exchange-rate adjustment process are subject to change over time.

7 Implications and Interpretations: An Evolving Standard?

Finally, what does our model say about the workings of the dollar-sterling exchange under the classical gold standard? What are its implications as compared with the traditional approach to gold point measurement? Table 5 and Figure 2 allow us to compare Officer's (1996) gold point estimates with our implied threshold values, all now measured as a percentage of par.

The comparison is revealing. As already noted, the behavioral thresholds

Table 5: The Gold Points as Measures of Integration: Direct Calculations and Estimated Decadal Thresholds Compared

year(s)	--- Officer's gold points --- (% of par, absolute)				Our threshold
	GPA-X	GPA-M	GTF-X	GTF-M	(% of par) gamma
1880-89	0.66	0.71	0.36	0.18	0.31
1890-99	0.66	0.63	0.54	0.35	0.46
1900-09	0.50	0.60	0.43	0.31	0.41
1910-13	0.50	0.59	0.43	0.36	0.15

Notes: GPA refers to Gold Point Arbitrage and GTF to gold transfer of funds. X and M refer to export and import gold points respectively. GPA is two-way pure arbitrage via an international swap of bills and gold; GTF is one-way payment of funds in gold, rather than in bills, to settle accounts for the shipment of other goods and services. See Officer (1996) for more details.

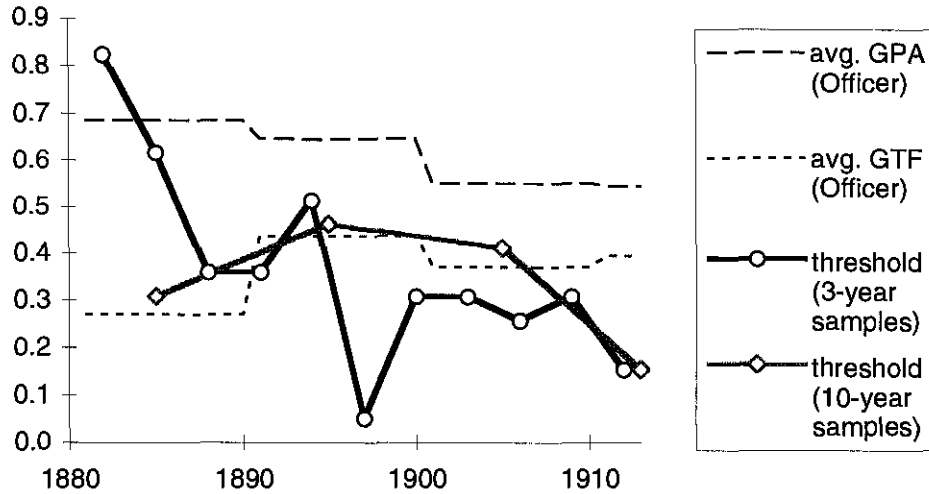
Sources: Officer (1996, 174) and Table 1.

show a marked evolution across time towards lower values; but the conventionally measured gold points showed no such tendency, and after 1880, proved remarkably stable. Our indirectly measured thresholds diminish in size by a factor of about four, but the directly measured gold points are stable or only slightly declining by at most one fifth. The difference can be seen late in the period in Table 5, where decadal samples are used. The trend is even clearer when we exploit our high-frequency data and estimate the thresholds every three years in Figure 2.¹⁵

The question now arises, what could account for these different trends in measured costs and implicit costs? It might be thought that this finding casts doubts on Officer's (1996) cost estimates, but we do not think so. Rather,

¹⁵We are also grateful to Lawrence Officer for the provision of (unpublished) monthly gold-point estimates for a continuous period 1890-1906. The coverage here is not as complete over time as in Figure 2 where we used Officer's decadal estimates. For comparison, in 1890 his monthly gold export point was about 0.53 on average, and fell to 0.47 by 1906, with some higher levels and volatility in between, around 0.6-0.7. This looks a lot like our estimated threshold, and leaves more questions about the uniformity of gold points over time. Unfortunately, Officer's *monthly* estimates do not extend to the 1880s or to the period 1907-13, the two episodes where we find significantly higher and lower thresholds than his direct gold point estimates suggest.

Figure 2: The Gold Points as Measures of Integration: Direct Calculations and Estimated Triennial Thresholds Compared



Source: Tables 3 and 5.

there may be “dynamic” rather than “static” sources of arbitrageurs’ cost-benefit calculations. We can think of many possible explanations. There are, perhaps, two observations to be explained. Why are our thresholds larger than the measured costs in the early period (1880s)? And why are our thresholds smaller than the measured costs in the later years (circa 1910)?

On the first question, a natural answer would seem to be the relatively immature stature of the gold standard in the 1880s. At that time, the U.S. had only recently made a commitment to gold following a resumption of convertibility in 1879 after many years on a paper currency during and after the Civil War, the so-called Greenback suspension. Thus, the gold basis for the dollar-sterling exchange was by no means credible or long-established, and reputational effects might have been weak. Such concerns might have reasonably affected the cost-benefit decisions of those willing to engage in arbitrage, until such times as the dollar-sterling exchange appeared safe and reliable. Later on, experience showed the volatility of the exchange was low, and not just temporarily so, perhaps encouraging more voluminous and

more efficient arbitrage at smaller deviations from parity. The tentative evidence that thresholds increased during economic and financial crises provides further evidence that arbitrage activity might have been sensitive to reputational and commitment concerns: our estimated thresholds increase during the height of the U.S. silver movement circa 1896 and, somewhat less dramatically, during and just after the 1907 financial panic.

At the other end, we can also conceive of reasons why, by around 1910, our threshold estimate appears so much lower than the gold point cost measure. An appealing explanation is simply the appearance of speculative activity, which is not captured in our model. In this way of thinking, the gold point band operates like an exchange-rate target zone (cf. Krugman 1991). Only here, the private speculators do not speculate against the government intervening at the edges of the band (as in modern settings under quasi-fixed exchange rates); rather, the speculation is against arbitrageurs (other private agents) intervening at the edge of the band. Thus, a gold standard could be seen to operate as a target zone with the intervention in “support” of a currency having been “privatized” and turned over to arbitraging agents, whilst the central bank merely fulfills a passive role as rule-driven buyer and seller of bullion, coin, and paper. In fact, exactly this kind of target-zone model has been suggested as the relevant model for the interwar gold-exchange standard, in a paper by Hallwood, MacDonald, and Marsh (1996).

Yet how could such a target-zone dynamic evolve? In the Hallwood et al. model, the exchange rate is everywhere driven by AR(1) dynamics, reflecting the within-gold-point speculative activity that prevents the exchange rate reaching the edge of the band. Speculators know that near the edge of a band the arbitrage possibilities give rise to almost-one-way bets: the exchange rate is much more likely to move toward the middle of the band than move further out. These dynamics are very different to the TAR model used here. The narrowing of our threshold spread is consistent with the emergence of a target-zone-like behavior (AR dynamics) during the late classical gold standard era, replacing our pure arbitrage system (TAR dynamics). This presaged the more widespread—and, some might say, destructive—speculation that reputedly resurfaced during the abortive interwar gold standard. In the algebra of our model, a smooth switch from TAR to AR dynamics is

achieved in only one way: by the gradual shrinking of the threshold value away from the gold-point level and down toward zero. This indeed appears to be happening late in the period.

Thus, our study offers several new issues for consideration. We have shown that work on the classical gold standard need not be confined to work on annual, quarterly, monthly, or even weekly data. Many basic financial publications list daily quotes for a wide variety of currencies, and we sample just one to construct a new database of dollar-sterling exchange rates over more than thirty years, 1879–1913. Such data are essential for the proper study of exchange rate dynamics where arbitrage operations were measured in days.

We then showed how such a study of gold arbitrage could be based on a simple model, and pointed out that price dynamics contained sufficient statistics to discern the nature of the underlying arbitrage technology. The model was applied using threshold autoregression methods, which yield estimates of a threshold separating two distinct regimes of exchange rate behavior: an inner band with a random walk, and an outer band where the exchange rate reverts inwards. Naturally, the thresholds have an interpretation as implicit or revealed-preference “gold points”; and, in theory, they do correspond to the minimum level of marginal costs where gold flows are induced.

We compared our estimated thresholds with accepted estimates of the gold points, and we found very different trends in the two measures. Our threshold declined dramatically from 1879 to 1913, but the measured gold points were fairly stable. We take this to be evidence that the classical gold standard was an evolving standard in ways not captured by pure costs measures alone. Our conjecture is that various dynamic considerations could have acted to cause arbitrage to operate at different thresholds: for example, evolving concerns over the reputation of the gold commitment, and the evolution of new technologies (or “learning”) in the market itself including, but not restricted to, the development of speculative activity and target-zone dynamics.

Thus, though the example *par excellence* of monetary stability in an international setting, the classical gold standard may yet deserve analysis as more than just a monolithic, rule-bound system driven by the simple arith-

metic of arbitrage operations. And it should be seen that the analysis of the gold points is more than an accounting exercise, and is a key ingredient in a fundamental method of market integration analysis, the comparison of measured costs to revealed behavioral responses. Our technique promises to be of use in all fields of market integration research. As for the gold standard, further studies will be needed to examine the workings of arbitrage in yet more microeconomic detail, and, following the present study, will need to exploit data at very high frequencies. Opportunities abound to extend this type of study, say, to other currencies, or to the brief interwar gold-exchange standard. However, our results suggest that there is life in this line of research yet, and even a few surprises.

Appendices

Appendix A: Tsay's Nonlinearity Test

A general specification test is provided by Tsay (1989) for TAR alternatives against an AR null. It is a simple nonparametric test for exactly the kind of nonlinear adjustment process we have in mind here. It has already been applied to the analysis of price processes in the study of arbitrage and market integration (Coleman 1995). The method is as follows. Following Tsay, our TAR model is of the general form

$$x_t = a_0^i + a_1^i x_{t-1} + e_t^i \quad \text{when } r(i) > x_{t-1} \geq r(i-1); \quad (9)$$

where the $r(i)$ for $i = 1, \dots, k-1$ are nontrivial threshold values, with $k \geq 2$, $r(0) = -\infty$, $r(k) = +\infty$. We call (x_t, x_{t-1}) a *case* of data for $t = 1, \dots, T$. We next order the cases according to the delayed level x_{t-1} , the lag of x that decides between which thresholds each case is located. Tsay's test is based on recursive residuals from the arranged AR(1) autoregression of this case data. The intuition is as follows. Suppose we order the cases in increasing fashion, and the first m cases fall below the first threshold. Then, the recursive AR(1) residuals will be orthogonal to the dependent variables of the regression. However, because of the regime change at the threshold level, the recursive residuals for cases after the m th will no longer have this same orthogonality. This suggests a simple F -test for the recursive residuals.

In full, let the ordered case data be (x_{t_s}, x_{t_s-1}) for $s = 1, \dots, T$. We perform an arranged AR(1) autoregression on this ordered data and generate recursive residuals e_{t_s} for each case. We then do an OLS regression of these predictive residuals on the dependent variables,

$$e_{t_s} = \omega_0 + \omega_1 x_{t_s-1} + u_{t_s}, \quad (10)$$

and the orthogonality test amounts to calculating the conventional associated F statistic for this regression, which follows an asymptotic F distribution for large T . One final item is to note that, especially in small samples, the case data may not fall in all of the regimes delineated by every threshold value. Thus, as a practical matter, the Tsay test should be run with both increasing

and decreasing ordering of the arranged regression. We follow that practice in what follows, and we report only the most significant p -values for each of the two F tests.

Appendix B: General TAR Estimation Method

This appendix describes the basic TAR estimation method used in the paper. The object is to model an equation of the form

$$\Delta x_t = \begin{cases} \kappa^{out} + \lambda^{out}(x_{t-1} - \gamma) + e_t^{out} & \text{when } x_{t-1} > \gamma; \\ \kappa^{in} + \lambda^{in}x_{t-1} + e_t^{in} & \text{when } \gamma \geq x_{t-1} \geq -\gamma; \\ \kappa^{out} + \lambda^{out}(x_{t-1} + \gamma) + e_t^{out} & \text{when } -\gamma > x_{t-1}. \end{cases} \quad (11)$$

where e_t^{out} is $N(0, \sigma^{out^2})$, e_t^{in} is $N(0, \sigma^{in^2})$, and we initially impose the restriction $\lambda^{in} = 0$.

The likelihood function for the above TAR model is

$$\begin{aligned} L_a &= L_a(\lambda^{out}, \lambda^{in}, \sigma^{out}, \sigma^{in}; \gamma) = \\ &= -\sum_{I_{in}(x_{t-1})=1} \frac{1}{2}(\log(2\pi) + \log(\sigma^{in^2}) + e_t^{in^2}/\sigma^{in^2}) \\ &= -\sum_{I_{out}(x_{t-1})=1} \frac{1}{2}(\log(2\pi) + \log(\sigma^{out^2}) + e_t^{out^2}/\sigma^{out^2}), \end{aligned} \quad (12)$$

where $I_{in}(x_{t-1}) = I(|x_{t-1}| \leq \gamma)$, and $I_{out}(x_{t-1}) = I(|x_{t-1}| > \gamma)$ are indicator functions which depend on the position of x_{t-1} being inside or outside the band.

The null AR(1) model is

$$\Delta x_t = \lambda x_{t-1} + e_t, \quad (13)$$

where e_t is $N(0, \sigma^2)$, with a likelihood function

$$L_n = -\sum_t \frac{1}{2}(\log(2\pi) + \log(\sigma^2) + e_t^2/\sigma^2). \quad (14)$$

We define the likelihood ratio as $LLR = 2(L_a - L_n)$, and our objective is to maximize this ratio. We follow a search algorithm:

1. Choose a set of candidate thresholds to search. We chose one quarter cent intervals from zero to ten cents. Set $k = 1$.
2. Choose γ equal to k th highest candidate threshold. Call this γ_k .

3. Partition the sample into observations inside and outside the band. Construct indicator values $I_{in}(x_{t-1})$ and $I_{out}(x_{t-1})$. If fewer than m values fall in each subsample, skip this value of k . We chose $m = 30$.

4. Calculate L_a as above, either by maximum likelihood, or, equivalently, by OLS on the partitioned samples. Then calculate $LLR = 2(L_a - L_n)$.

5. Increase k by one and go to step 2.

6. Locate the choice of k and γ_k that maximizes LLR . Note the associated $(\lambda^{out}, \lambda^{in}, \sigma^{out}, \sigma^{in})$ estimated at step 4 for that choice of k . This yields a maximum likelihood estimate of $(\lambda^{out}, \lambda^{in}, \sigma^{out}, \sigma^{in}, \gamma)$.

Since γ is not identified under the null, standard inference is invalid, and LLR does not follow the usual χ^2 distribution of likelihood ratio tests. An alternative is to follow Monte Carlo simulation methods. We proceed as follows:

1. Estimate the AR(1) null model (13) on the actual data (x_1, \dots, x_T) .

2. Generate $i = 1, \dots, n$ simulations of this model. We chose $n = 1,000$. Start each with $x_{-b} = 0$, end at x_T , and discard the first b values x_{-b}, \dots, x_{-1} to avoid initial value bias. We chose $b = 50$.

3. For this i th simulation, estimate the TAR model as above and calculate the simulated log likelihood ratio LLR_i .

4. Calculate the empirical distribution of the LLR_i , and use this as the basis for inference in judging the alternative TAR model (11) against the AR(1) null (13).

References

- Balke, N. S., and T. B. Fomby. "Threshold Cointegration." *International Economic Review* (1997). Forthcoming.
- Bloomfield, A. I. *Monetary Policy Under the International Gold Standard: 1880–1914*. New York: Federal Reserve Bank of New York, 1959.
- Clark, T. A. "Violations of the Gold Points, 1890–1908." *Journal of Political Economy* 92 (October 1984): 791–823.
- Cole, A. H. "Evolution of the Foreign-Exchange Market of the United States." *Journal of Economic and Business History* 1 (May 1929): 348–421.
- Coleman, A. M. G. "Arbitrage, Storage, and the Law of One Price: New Theory for the Time Series Analysis of an Old Problem." Princeton, N.J., September 1995. Photocopy.
- Davis, L. E., and J. R. T. Hughes. "A Dollar-Sterling Exchange, 1803–1895." *Economic History Review* 13 (August 1960): 52–78.
- Eichengreen, B. J. "Trends and Cycles in Foreign Lending." In *Capital Flows in the World Economy*, edited by H. Siebert. Tübingen: Mohr, 1990.
- Einzig, P. *The History of Foreign Exchange*. 2nd ed. London: Macmillan, 1970.
- Fanizza, D. G. "Multiple Steady States and Coordination Failures in Search Equilibrium: New Approaches to the Business Cycle." Ph.D. dissertation, Northwestern University, 1990.
- Goodhart, C. A. E. *The New York Money Market and the Finance of Trade, 1900–1913*. Cambridge: Harvard University Press, 1969.
- Goschen, G. J. *The Theory of the Foreign Exchanges*. London: Effingham Wilson, 1861.
- Hallwood, C. P., R. MacDonald, and I. W. Marsh. "Credibility and Fundamentals: Were the Classical and Interwar Gold Standards Well-Behaved

Target Zones?" In *Modern Perspectives on the Gold Standard*, edited by T. Bayoumi, B. J. Eichengreen and M. P. Taylor. Cambridge: Cambridge University Press, 1996.

Krugman, P. R. "Target Zones and Exchange Rate Dynamics." *Quarterly Journal of Economics* 106 (August 1991): 669–682.

Moggridge, D. E. *British Monetary Policy 1924–1931: The Norman Conquest of \$4.86*. Cambridge: Cambridge University Press, 1972.

Morgenstern, O. "The Validity of International Gold Movement Statistics." *Princeton Special Papers in International Economics* (1955).

Morgenstern, O. *International Financial Transactions and Business Cycles*. Princeton: Princeton University Press, 1959.

Newcomb, S. *Principles of Political Economy*. New York: Harper & Brothers, 1886.

Obstfeld, M. "International Capital Mobility in the 1990s." In *Understanding Interdependence: The Macroeconomics of the Open Economy*, edited by P. B. Kenen. Princeton, N.J.: Princeton University Press, 1995.

Obstfeld, M., and A. M. Taylor. "Nonlinear Aspects of Goods-Market Arbitrage and Adjustment: Heckscher's Commodity Points Revisited." *Journal of the Japanese and International Economies* (1997a). Forthcoming.

Obstfeld, M., and A. M. Taylor. "The Great Depression as a Watershed: International Capital Mobility in the Long Run." In *The Defining Moment: The Great Depression and the American Economy in the Twentieth Century*, edited by M. D. Bordo, C. D. Goldin and E. N. White. Chicago: University of Chicago Press, 1997b. Forthcoming.

Officer, L. H. "Dollar-Sterling Mint Parity and Exchange Rates, 1791–1834." *Journal of Economic History* 43 (September 1983): 579–616.

Officer, L. H. "Integration in the American Foreign-Exchange Market, 1791–1900." *Journal of Economic History* 45 (September 1985): 557–85.

- Officer, L. H. "The Efficiency of the Dollar-Sterling Gold Standard, 1890–1908." *Journal of Political Economy* 94 (October 1986): 1038–73.
- Officer, L. H. "The Remarkable Efficiency of the Dollar-Sterling Gold Standard, 1890–1906." *Journal of Economic History* 49 (March 1989): 1–41.
- Officer, L. H. "Gold-Point Arbitrage and Uncovered Interest Arbitrage Under the 1925–31 Dollar-Sterling Gold Standard." *Explorations in Economic History* 30 (January 1993): 98–127.
- Officer, L. H. *Between the Dollar-Sterling Gold Points: Exchange Rates, Parity, and Market Behavior*. Cambridge: Cambridge University Press, 1996.
- Perkins, E. J. *Financing Anglo-American Trade*. Cambridge: Harvard University Press, 1975.
- Prakash, G. "Pace of Market Integration." Northwestern University, September 1996. Photocopy.
- Seyd, E. *Bullion and Foreign Exchanges*. London: Effingham Wilson, 1868.
- Spalding, W. F. *Foreign Exchange and Foreign Bills*. London: Pitman, 1915.
- Spiller, P. T., and R. O. Wood. "Arbitrage During the Dollar-Sterling Gold Standard 1899–1908: An Econometric Approach." *Journal of Political Economy* 96 (August 1988): 882–92.
- Tsay, R. S. "Testing and Modeling Threshold Autoregressive Processes." *Journal of the American Statistical Association* 84 (March 1989): 231–240.
- Zevin, R. B. "Are World Financial Markets More Open? If So, Why and With What Effects?" In *Financial Openness and National Autonomy: Opportunities and Constraints*, edited by T. Banuri and J. B. Schor. Oxford: Clarendon Press, 1992.