NBER WORKING PAPER SERIES

MEASURING MARKET INTEGRATION: FOREIGN EXCHANGE ARBITRAGE AND THE GOLD STANDARD, 1879-1913

Eugene Canjels
Gauri Prakash-Canjels
Alan M. Taylor

Working Paper 10583
http://www.nber.org/papers/w10583

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2004

Forthcoming in *The Review of Economics and Statistics*. We thank Lawrence Officer for helpful conversations. We thank Andrew Coleman for useful feedback and for permission to use his data on gold flows. We are grateful to Nathan Balke, Martin Bayerle, Jan Tore Klovland, Paul O'Connell, Maurice Obstfeld, Christopher Udry, Claire Waters, Mark Watson, the editor, and an anonymous referee for their comments, and we are similarly indebted to 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. The views expressed herein are those of the author(s) and not necessarily those of the National Bureau of Economic Research.

Measuring Market Integration: Foreign Exchange Arbitrage and the Gold Standard, 1879-1913
Eugene Canjels, Gauri Prakash-Canjels, and Alan M. Taylor
NBER Working Paper No. 10583
June 2004
JEL No. N1, F3

ABSTRACT

A major question in the literature on the classical gold standard concerns the efficiency of international arbitrage. Authors have examined efficiency by looking at the spread of the gold points, gold point violations, the flow of gold, 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 nonlinear theoretical models and threshold autoregressions. We also compile a new, high-frequency series of continuous daily data from 1879 to 1913. We can derive reasonable econometric estimates of the implied gold points and price dynamics. The changes in these measures over time provide an insight into the evolution of market integration.

Eugene Canjels
Economic Consulting
Deloitte & Touche LLP
555 12th Street, NW
Washington, DC 20004
ecanjels@deloitte.com

Gauri Prakash-Canjels
The CapAnalysis Group
Howrey, Simon, Arnold & White
1299 Pennsylvania Avenue, NW
Washington, DC 20004
prakash-canjelsg@capanalysis.com

Alan M.Taylor
Department of Economics
University of California
One Shields Avenue
Davis, CA 95616
and NBER
amtaylor@ucdavis.edu

1 Introduction

How can we measure market integration? By way of illustrating a new approach to the problem, this paper assesses the degree of market integration in the dollar-sterling foreign exchange market of 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. A major issue has been whether the gold-standard regime was stable and efficient, and tests of this have often centered on whether the gold points bounded the market exchange rate. The gold points, the exchange rates at which gold arbitrage would commence, then needed to be estimated to facilitate the test. Thus, the literature has frequently focused on getting this estimation right. However, a persistent lacuna in this literature is suggested by the disconnect between the arduous work devoted to measuring the gold points, and the apparent lack of progress in documenting the connection between these measures and the actual behavior of arbitrageurs in the market.

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: 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 movements of the exchange rate, even in monthly time series.[1]

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 finds much at fault in the revisionist literature. For example, the major studies all incorrectly used the cable transfer rates of exchange when the dominant arbitrage instrument was the demand bill (or sight bill), and, most importantly, they used *ad hoc* gold point estimates based on secondary sources. Officer's response was to meticulously recalculate gold points

[1] This approach was revisited by Clark (1984), who reached similar conclusions using refined estimates of transactions costs and found, 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).

Table 1: Dollar-Sterling Exchange Rate, Officer's Monthly Data, 1791–1966

	Percent sterling premium over parity				
Period	Mean	Mean absolute	Std. dev. about mean	Std. dev. about zero	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, p. 103).

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. By the late nineteenth century, the dollar-sterling exchange had reached its peak level of stability, 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 match levels seen under the classical gold standard (Eichengreen 1991; Obstfeld and Taylor 2004).

We believe that significant new contributions to the literature can be made in two ways. First, we can better formalize the theory of arbitrage in the markets concerned; and, second, we can 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]

One deficiency in most of this literature, identified by Officer (1996, p. 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. A major aim of this paper is to formally model the linkage between the two for the first time.

Officer (1996, pp. 117–21) arrived at his method by discounting all the others. What are those alternative methods? Officer's nine categories can be collapsed into four:

> **(a) Consult an Expert or Two.** The common method employed by newspapers a century ago. An "expert" typically meant a foreign-exchange dealer or a banker. But such sources may be inconsistent, unrepresentative, omit some costs, and may not provide a complete time series.
> **(b) Find the Exchange Rate at Which Gold Flows.** This elegant method has obvious theoretical appeal, using as it does the principle of revealed preference. It was suggested over one hundred years ago by Newcomb (1886, pp. 281–82), but his suggestions were neglected. The method has been applied by various authors (e.g., Clark 1984), but has run into practical difficulties due to the poor quality of gold-flow data.[3] However, later in the paper we report some new, seemingly reliable, and previously unused data on gold flows that re-open this possibility.
> **(c) Use the Exchange-Rate Maximum and Minimum.** The advantage of this technique is the minimal data requirement. The main problem is the maintained assumption: that any gold-point violations are eradicated so instantaneously by arbitrage that we never observe them. This is clearly, a priori, a tenuous assumption; in a moment we will show that it is in fact erroneous.
> **(d) Break Down All Components of Transaction Costs.** The benefit is that it makes no maintained theoretical assumptions. The obvious problem is the intense data requirement. A less-obvious problem is whether these direct cost estimates do indeed correspond to the behavior of the arbitrageurs in question.

Officer proved the feasibility of method (d) whilst casting grave doubt on

[2] See, *inter alia* Moggridge (1972), Davis and Hughes (1960), Bloomfield (1959), Morgenstern (1959), Cole (1929), Spalding (1915), Seyd (1868), Goschen (1861), and Officer (1996).

[3] 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.

method (a).[4] Still, it remains an open question whether the methods (b) and (c) can be improved upon, and whether they might shed light on the actual process of arbitrage, one of the weaknesses of method (d). A major contribution of this paper is to show how to implement a novel variant of Newcomb's elegant test in method (b) using a nonlinear model to detect the points at which the onset of arbitrage behavior begins, whilst 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 in method (c). Even so, for the few years in which new sources of data can supply more reliable gold-flow data we can perform a cross-check, where we find qualitative evidence to support our conclusions.

If our approach is to be feasible, however, a richer dataset is needed than has been hitherto assembled. Accordingly, in the final and most labor-intensive contribution of the paper, we have constructed by hand a new dataset of high-frequency (daily) exchange rates from newspapers, a new time-series with a frequency that more closely corresponds to the adjustment horizon in the actual market. We next discuss these data, and the operation of the market, before moving on to the theory and empirics.

2 Data

2.1 Exchange Rates

Previous researchers have relied on monthly, quarterly, or annual data series. Although 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 every full calendar year of the dollar-sterling gold standard that began with the restoration of U.S. convertibility after the Greenback suspension on January 1, 1879, and ended with the suspension of U.K. convertibility in August 1914 (Officer 1996, pp. 16 and 43).

We collected the exchange rates based on sterling demand bills in New York since it is now accepted that the dominant form of arbitrage under the classical gold

[4]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 of Clark (1984) that assume some components of costs, or use cost aggregations, are properly tested against the strict and unforgiving ruler of repeated fine-scale measurement.

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 sixty-day bill introduced in colonial times was the preferred instrument, except for a brief period in the mid-nineteenth century when a three-day bill was used (Officer 1996, pp. 113–15).[5]

We went to the best primary source for these data, the *Financial Review*, which tabulated daily data on the New York exchange each year in an annual summary.[6] Sight bill quotations are usually in a range, e.g., "4-85 85-1/2" meaning between 4.85 and 4.855 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.[7] 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 cover 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 1879 to 1913. This represents an enormous amount of data for time-series analysis: over 10,000 observations in the entire sample. With such a dataset we should have ample information to identify parameters of interest and how they change over time.

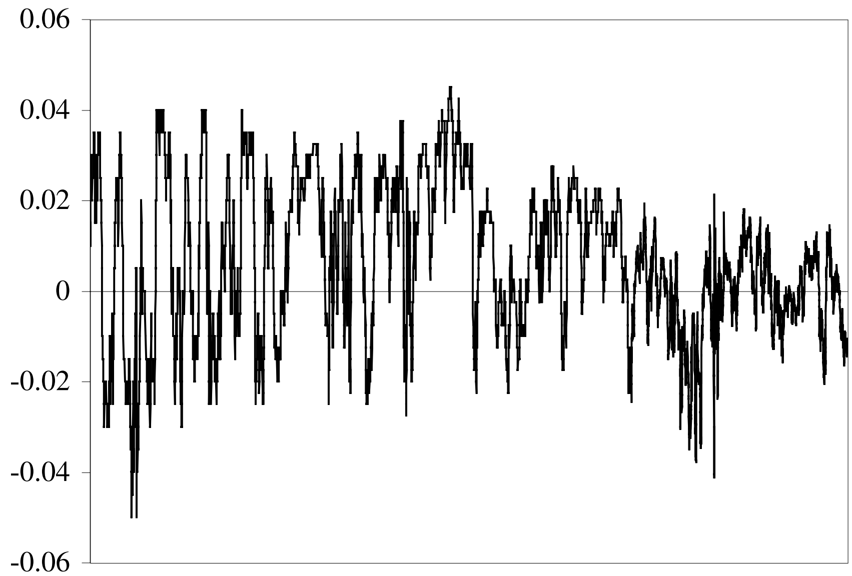
Figure 1 displays the exchange rate deviation $x_t = E_t - E^{\text{par}}$ measured in dollars where we define parity as the ratio of the fixed mint prices of gold in the United States (P^g, in dollars per ounce) and Britain (P^{g*}, in pounds per ounce),

[5] 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 (Davis and Hughes 1960; Perkins 1975; Officer 1996, pp. 60–63 and 115).

[6] We 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, and this proved beyond our scope. Instead, we took the thirty five annual summaries and entered the daily data on "posted rates."

[7] Specifically, before 1904 the finest distinction is 0.5 cents, thereafter 0.05 cents.

Figure 1: Dollar-Sterling Exchange Rate, Daily Data, 1879–1913



Note: The figure shows the deviation from parity, $E^{\text{par}} = 4.86656$, in dollars.
Source: *Financial Review*.

that is, $E^{\text{par}} = P^g/P^{g*} = 4.86656$. We note that the maximum deviation from parity over the full period was slightly more than \$0.05 or about 1.06%. The dynamics of x_t will be the object of study in the remainder of the paper. We think it is fairly clear that x_t does not exhibit explosive behavior and we will assume stationarity in all inference and seek to identify any nonlinear dynamics in the series as suggested by theory.[8]

2.2 Gold Flows

Though our initial focus is mainly on arbitrage as it relates to price dynamics, we will later perform some cross-checks on our analysis by using two new sources of gold flow data. The quality of gold flow data was first seriously questioned by

[8] If it is not obvious from the chart, we can report that the Augmented Dickey Fuller test statistic (with intercept, linear time trend and three lags chosen by the Schwarz Criterion) was a highly significant -7.22.

Morgenstern (1955), who found serious discrepancies in many periods between the reported imports of gold received and the exports of gold sent for several countries, including the U.S. and Britain. The source of Morgenstern's data was the Monthly Summary of Commerce and Finance of the United States published by the Department of Commerce. Goodhart (1969) performed a much-needed correction for the period 1900–13 to strip out various mistakes, and he found that there was then a fair match between export and import data provided by the U.S. and Britain. However, Goodhart's data are available only on a monthly basis, which is too low a frequency to be of use in a model of arbitrage activity.

Prospecting for new gold flow data, we made a lucky strike. After writing most of this paper, we discovered previously unused (at least for this purpose) gold export data published in the Annual Reports of the Director of the U.S. Mint. Between May 1888 and July 1889 about $60 million in gold was exported from the U.S., and this was a matter of great concern for the Director of the Mint. In response he began publishing tables in his annual report documenting the exact date, quantity, and destination of gold exports from New York, which was the main port of departure. The data are apparently directly obtained by the Assay Office in New York from Customs reports and appear to be of better quality than any previous data put to use, although, as with most gold export data, there seems to be an occasional problem of figuring out the precise destination of any shipment.

We compiled these daily data from the original source as soon as they came to our attention—which was thanks to Captain Martin Bayerle, who had examined this source for quite a different purpose, in an effort to prove that the liner *RMS Republic*, which sank in fog two days after sailing for France from New York on January 22, 1909, was carrying a large and clandestine gold shipment.[9] He writes: "I found the French import data very reliable, and an almost exact match between US exports and French imports exists for the years 1904–14" (personal communication). In many cases there is an almost exact match between these figures and Goodhart's numbers, although on several occasions there are large discrepancies.

[9]Thus, the loss of the *Republic*, even at the time the biggest maritime loss in history, could now precipitate the most lucrative salvage operation ever known. Bayerle believes the cargo was an enormous quantity of bullion and coin intended by the U.S. government as a subscription to part of the massive Russian bond issue of that year, making its movement—and perhaps all the more so its irrecoverable loss—a very politically sensitive matter. Conspiracy theorists have long stoked these rumors, egged on by the official silence of the American authorities and the unprecedented failure of the British Board of Trade, under Winston Churchill, to hold the requisite legal enquiry, the findings of which might have averted the loss of another White Star vessel, the *RMS Titanic*, three years later. See <http://www.rms-republic.com>.

Although, one has to be fairly skeptical about any one particular observation of gold exports, and possibly some gold exports are missed in these tables, overall they seem sufficiently detailed and reliable as a check on our methods.

Sadly, no comparable source for gold imports exists as the director of the Mint was less worried about heavy gold inflows. The best data that we are aware of was compiled and generously provided by Andrew Coleman (1998). He collected weekly shipping reports published in the New York Times between March 1895 and November 1901. The quality of these data is also an issue, as newspaper reports are considered to be the least reliable source, and the short time span is also a major disadvantage. Unfortunately, these data provide only weekly aggregates and only for a limited number of years.

3 A Model of Gold Point Arbitrage

We next explore the workings of the market and construct a tractable model of gold point arbitrage based on certain key features: the demand for sterling sight bills in New York, changes in the net supply of bills via gold point arbitrage, and the costs and benefits of the arbitrage operation itself.

3.1 Demand for Sterling Bills in New York

Leonard Presnell, perhaps only half jokingly, once declared that the "international gold standard" was a misnomer, and that the regime would be better described by the term "the international bill-on-London standard" (Davis and Gallman 2001, 131). The serious point here is that the functioning of the sterling-centered system depended not only on the free convertibility of gold both in London and overseas, it also involved the ability to quickly and easily translate domestic currency claims into sterling claims, and vice versa, using an important "quasi-money"—the legendary financial instrument knows as the sterling demand (or sight) bill. Innovated in centuries past this was a negotiable instrument that was almost universally acceptable in trade and finance circles, and hence highly liquid. It could serve as a means of payment for trade, or, if remitted to London and cashed on sight (that is, on demand), it could perform as a vehicle for capital movement. Accordingly we think of the demand for a stock of sterling sight bills in offshore centers, such as New York, as being analogous to a demand for quasi-money, and we will base our analysis on a partial equilibrium model of that market.

Let the stock of bills in the New York market be B_t.[10] 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,

$$E_t = \psi - \eta B_t + u_t, \tag{1}$$

where ψ 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.[11]

Both gold and sterling bills will circulate in the model between two centers, London and New York. We note at this point that although there are two centers and two goods in the model, meaning four prices, we do not need to consider the market for bills in London, nor the market for gold in both centers, since in those markets we can effectively treat prices as fixed, and all arbitrage in this system is driven by one price, that of sterling demand bills in New York. The market price of gold in each center was fixed by the mints at the parity level adjusted for the relatively fixed transaction costs of buying and selling.[12] The market price of sterling bills in London, like those of a check to be cashed, was equal to their face value plus or minus similar transaction costs (that is, a one pound bill in London was worth, effectively, one pound). These institutional features motivate our approach of considering a very simple form of price adjustment in E_t only, and argue against the application of a commodity-market type of model with price adjustment in both locations (Coleman 1998).

The time series disturbance term 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

[10]The extent of this market could, of course, include the entire United States, much of it linked to the New York market. The development of the internal U.S. market and its relationship to the external market has been a subject of considerable debate (Davis and Hughes 1960; Officer 1996).

[11]In 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.

[12]However, the model can be easily extended to apply to "triangular" arbitrage via third markets and to situations where the gold price in one or more centers is subject to exogenous variation. Coleman (2002) discusses arbitrage between New York and London via Paris and cases where the London gold price was affected by the use of gold devices by the Bank of England.

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, or the desire of agents to make shifts in their dollar versus sterling portfolios. We consider all such derived demand, supply, and "taste" shocks as exogenous shifts and for the present purpose we impound all these effects in the disturbance term.

3.2 The Mechanics of Gold Point Arbitrage

We next consider how gold and sterling bills circulate between the two centers. 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. 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, $E^{\text{par}} = P^g / P^{g*}$. Provided such revenues exceeded transaction costs, the trade would be profitable.

Table 2 relates the changes in the quantity of gold and bills in the typical transactions. The table shows first that there was no delay between the arbitrageurs' actions and the change in the stock of bills in the New York market and this implies that we can ignore bill shipment delays. It is never the case that the New York market has to wait for bills to arrive from another center like commodities—they are simply created by financial intermediaries.[13] The table shows that, in both directions, the relationship between gold movements and changes in the stock of demand bills is given by

$$\Delta B_t = -P^{g*} \Delta G_t, \tag{2}$$

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

A remarkable feature of gold point arbitrage was that there was little or no risk in the realization of profit. Suppose that the New York arbitrageur takes profits in dollars in New York. In the case of gold export the revenue is taken out before gold is acquired for shipment, and there is no risk, no interest cost, and profit is

[13] In contrast, Coleman's (1998) model includes shipment delays as would occur in commodity markets.

Table 2: Gold Import and Export Via Demand Bill

Time	Place	Action by arbitrageur	Gain	Loss
Export of one ounce of gold, with $E > E^{par}$				
$t = 0$	New York	sell demand bills	\$ EP^{g*}	bills £P^{g*}
$t = 0$	New York	buy gold	gold oz. 1	\$ $E^{\text{par}}P^{g*}$
$t = T$	London	sell gold	£P^{g*}	gold oz. 1
$t = T$	London	redeem bill	bills £P^{g*}	£P^{g*}
		Marginal revenue	\$ $(E - E^{\text{par}})P^{g*}$	
		Bills inflow	£$+P^{g*}$	
		Gold inflow	oz. -1	
		Interest cost in time	zero	
Import of one ounce of gold, with $E < E^{par}$				
$t = 0$	New York	buy demand bills	bills £P^{g*}	\$ EP^{g*}
$t = T$	London	redeem bill	£P^{g*}	bills £P^{g*}
$t = T$	London	buy gold	gold oz. 1	£P^{g*}
$t = 2T$	New York	sell gold	\$ $E^{\text{par}}P^{g*}$	gold oz. 1
		Marginal revenue	\$ $(E^{\text{par}} - E)P^{g*}$	
		Bills inflow	£$-P^{g*}$	
		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, pp. 111–13).

immediately realized.[14] 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 stochastic time of expected length T.[15] This introduces the possibility of an asymmetry in costs, and hence in the gold points. Still, the arbitrageur could engage in advance the shipping contracts, for a known price, and calculate expected profit, adjusting for any risk aversion to the stochastic time delay of two voyages. In neither case, however, was there uncertainty over prices, absent any default risk and neglecting any risk of a loss in transit, say to due to the sinking of a ship at sea bearing bills or gold. It is these unusual features of gold point arbitrage that make a simple, tractable model both possible and desirable.

[14]That 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 $t = 0$.

[15]With 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 $2T$.

3.3 Costs of Arbitrage

The final element of our model is the cost function for the arbitrage operation. We will consider the cost function to be a convex (for simplicity, quadratic) function of the flow of gold ΔG_t. We will consider a representative arbitrageur with a cost function for transactions that depends on the quantity of transactions. 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 $|\Delta G_t|$. We suppose that the total cost of the flow is given by

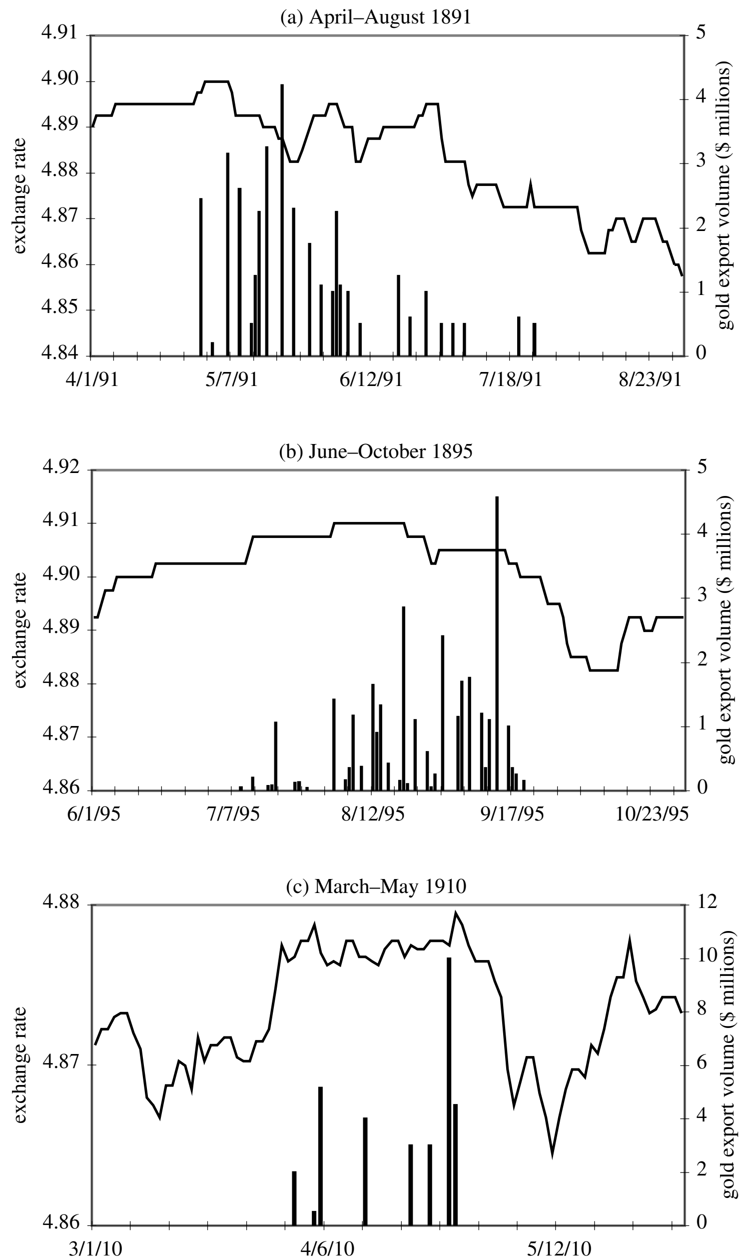
$$TC = b|\Delta G_t| + \frac{1}{2}c|\Delta G_t|^2. \tag{3}$$

This is a general technology where there are *no fixed costs*, 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 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. In the empirical analysis we will make explicit allowance for such asymmetries.

The arbitrage technology represents a departure from most of the traditional gold point literature which assumes, at least implicitly, constant marginal costs, with $b > 0$ and $c = 0$. From Table 2 we know that the marginal revenue of one extra ounce of gold movement is given by $MR = |E_t - E^{\text{par}}|P^{g*}$ Hence, in this traditional view, once the marginal revenue of shipment exceeds marginal cost b, gold freely moves (in the appropriate direction) and the exchange rate cannot move any further from parity. It is this view of the market that has motivated the use of the exchange rate maximum and minimum as the estimates of the gold points as in method (c). Upon closer inspection, we think that this view of the market cannot be maintained. Consider the following three periods of large gold export illustrated in Figure 2.

The first period of heavy gold exports was from April 29, 1891 to July 6, 1891; the second period was from July 9, 1895, to September 20, 1895; and the third period was from April 1, 1910, to April 26, 1910. In each of these episodes \$30 to \$35 million was exported from New York to England. As can be seen, in the beginning of May 1891 the exchange rate reached \$4.9 but exports still took place in the end of June when the exchange rate was only \$4.89. In the first half of October 1895 the exchange rate reached \$4.91, but again many exports took place in the beginning of September when the exchange rate did not go higher than

Figure 2: The Exchange Rate and Gold Exports, Daily, Three Episodes



Sources: *Financial Review* and U.S. Mint Reports.

$4.905. On April 26, 1910, the exchange rate reached $4.8795 but the first exports started when the exchange rate had not gone beyond $4.8775. Even if we allow for some errors in the gold export data it is hard to believe that these conclusions can be overturned and we conclude, in accord with Officer, that using the maximum and minimum observed price (method (c)) is not satisfactory as a way to estimate gold points.

We think a major reason for these results is that the marginal cost of gold point arbitrage increased with the quantity shipped. This had at least two causes.

Firstly, the arbitrage firms had a limited amount of capital available for their operations and using it for GPA made it unavailable for other uses. If they had a portfolio choice of where to invest their capital they would have first diverted it from the lowest yielding alternative opportunities, and later from investment opportunities with higher yields. This mechanism is obviously outside the realm of our partial equilibrium model, where we have purposely abstracted from portfolio choices. The second reason is that costs of shipping gold could go up when quantities increase, and this more directly fits our model. Moreover, there is copious evidence from contemporary reports to support this notion. For instance, gold could be exported either as gold bars or gold coins, where the first one was the least costly method for arbitrageurs.[16] However, when exports were heavy the U.S. Mint could not always provide gold bars, and arbitrageurs would have to resort to the more costly method of gold coins.

For example, on December 22, 1908, of a $500,000 shipment by Goldman Sachs the *Wall Street Journal* noted that “this engagement of gold bars represents the accumulation of daily receipts at the Assay Office during the past ten days, or since the National City Bank took all the suitable gold bars there when it shipped $4,000,000 on Dec. 12. The Assay Office is thus again without any exportable bars.” Such events were not unusual. On May 18, 1909, the *New York Post* stated that: “The demand for exchange was so much in excess of supply as to make exports extremely profitable with sterling bills selling at today’s high level. There was a rush to secure gold for Europe almost as soon as the market opened, and the fact that the coin obtainable at the Sub-Treasury was in very good condition led many bankers to risk the usual chance of abrasion in sending coin instead of bars.” Two days later the *Journal of Commerce* described a “scramble” for gold bars wherein banks were reserving ahead each day’s entire meltings and the Superintendent was

[16]Why? Compared to bars, coins would be bought at a discount by an overseas mint; they were more awkward to transport, being more easily abraded; and they were much easier to purloin (and one assumes, therefore, more expensive to insure).

forced to place limits on such tactics.[17]

It also happened at least once that the Assay Office of New York ran out of gold and that arbitrageurs had to secure gold from other Assay Offices with the additional cost of shipping the gold to New York. This caused considerable consternation, as the *New York Commercial* noted on January 12, 1909 under the headline "GOLD EXPORTS STOPPED BY LACK OF GOLD BARS: BANKERS EVEN GO TO PHILADELPHIA TO GET THEM":

> While it would be expensive to transport gold bars from Philadelphia here, international bankers attempted to secure bars yesterday rather than ship gold coin.…There has been much disapproval expressed against the small supply of gold bars at the Assay Office as on every occasion of gold exports last year the movement was stopped by this lack."

The next day large shipments were reported in the same newspaper "almost all in gold coin, a transaction that was not thought possible at any reasonable profit." As might be suspected, when exports were heavy the U.S. Mint often started to discourage exports through additional costs or inconveniences, and this is hinted at above. Thus government policy, by the deployment of such "gold devices," could have thrown sand in the wheels of gold point arbitrage, leading just as surely to increasing marginal costs.

We think the case for constant marginal costs cannot really hold in the face of this evidence. To model increasing costs we henceforth assume a convex cost function and in order to arrive at a closed form solution we restrict it to be quadratic, which is always valid as a first approximation.

3.4 Optimal Arbitrage

We now consider arbitrage operations. Arbitrageurs move gold up to the point where marginal revenue equals marginal cost. Without loss of generality, consider gold export at time t, where the import case is analogous. For export, we must have $E_t > E^{\text{par}}$ and $z > 0$ units of gold are already being moved. From Table 2 we know that the marginal revenue of one extra ounce of gold movement is $MR = (E_t - E^{\text{par}})P^{g*}$; and, from equation (3), the marginal cost is $MC = b + cz$. By equating MC and MR, and by a similar calculation for gold import, we obtain the

[17] Quotes here and below are as cited at <http://www.rms-republic.com>.

optimal flows $z = G_t - G_{t-1} = \Delta G_t$ as a function of exchange rate deviations $E_t - E^{\text{par}}$,

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

From equations (1) and (2), gold flows and the exchange rate are related via

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

where $v_t = \Delta u_t$. We think v_t, as the difference of u_t, is likely to be a stationary process. In what follows, we will assume that v_t is independent but not necessarily identically distributed.

A solution follows from equations (4) and (5). To simplify the model, let us normalize by defining $x_t = E_t - E^{\text{par}}$, so that x_t is the deviation of the exchange rate from par. We find in the gold export regime that

$$\begin{aligned} \Delta x_t &= \Delta E_t = \eta P^{g*} \Delta G_t + v_t \\ &= -\eta P^{g*} \frac{(E_t - E^{par})P^{g*} - b}{c} + v_t \\ &= -P^{g*} \alpha x_t + \alpha b + v_t. \end{aligned}$$

where $\alpha = \eta P^{g*}/c$. After some tedious manipulations of this equation we can recover the principal object of our study, the difference equation governing the dynamics of the exchange rate, as follows

$$\begin{aligned} (1 + \alpha P^{g*})x_t &= x_{t-1} + \alpha b + v_t; \\ x_t &= \frac{1}{1 + \alpha P^{g*}} x_{t-1} + \frac{\alpha}{1 + \alpha P^{g*}} b + \frac{1}{1 + \alpha P^{g*}} v_t; \\ \Delta x_t &= -\lambda(x_{t-1} - \gamma) + \mu v_t; \end{aligned} \tag{6}$$

where $0 < \lambda = \frac{\alpha P^{g*}}{1+\alpha P^{g*}} < 1$, $\gamma = \frac{b}{P^{g*}}$, and $\mu = \frac{1}{1+\alpha P^{g*}} < 1$. A similar derivation holds for the gold import regime, and in practice, of course, the model parameters might vary across regimes. When no arbitrage is profitable, the stocks of bills and gold remain unchanged and the exchange rate is driven by the disturbance term in equation (5). The complete solution of the model is then as follows:

$$\Delta x_t = \begin{cases} -\lambda(x_{t-1} - \gamma) + \mu v_t & \text{when } x_{t-1} > \gamma; \\ v_t & \text{when } \gamma \geq x_{t-1} \geq -\gamma; \\ -\lambda(x_{t-1} + \gamma) + \mu v_t & \text{when } -\gamma > x_{t-1}, \end{cases} \tag{7}$$

This model incorporates a simple formulation of gold point dynamics, a threshold autoregressive (TAR) model with three regimes and heteroskedasticity. If the current exchange rate is in the upper and lower regimes, then the exchange rate reverts toward the edge of the band (the nearest gold point) at a speed λ, where necessarily $0 < \lambda < 1$. Within the band, in the middle regime, between the gold points, there is no reversion and the process follows a driftless random walk. The gold-point spread γ and the adjustment speed λ are intuitively related to the shape of the underlying arbitrage cost function. An increase in the linear cost parameter b causes an increase in the gold point spread γ, as in the traditional approach; an increase in the quadratic cost parameter c causes a decrease in the speed of convergence λ. We also see that $\lambda = 1$ and adjustment takes only one period in the case where $c = 0$ and marginal costs are constant. This is intuitively obvious: if marginal costs never rise, then enough gold will be shipped right away to force the exchange rate back to the gold points. As we have argued, this case seems empirically less relevant, as such shipments are not seen, nor can costs be assumed to be linear forever.

4 The Econometric Model

The models developed in the theoretical section find their closest discrete-time econometric representation in the Threshold Autoregression model (or TAR; sometimes referred to as SETAR or Self-Exciting TAR). In such models the dynamics are governed by AR processes that differ across regimes delineated by the position of a lagged value of the dependent variable relative to a set of given thresholds (See Tong, 1983 and 1990, and Potter, 1999, for an overview).

The general form of a TAR model with R regimes, an order k, and a delay parameter d, can be expressed as

$$x_t = \beta_0^r + \sum_{i=1}^{k} \beta_i^r x_{t-i} + \epsilon_t^r, \text{ if } \gamma_{r-1} \leq x_{t-d} < \gamma_r$$

where $-\infty = r_0 < r_1 < ... < r_R = \infty$ and ϵ_t^r is a martingale difference sequence. In this model AR(k) dynamics obtain in regime r. The process is in regime r at time t when the selection variable x_{t-d} lies between two thresholds $\gamma_{r-1} < x_{t-d} \leq \gamma_r$, with $r = 1, \ldots, R$ and, by convention, $\gamma_0 = -\infty$ and $\gamma_R = +\infty$. We will call this model a TAR(R; k, d), or sometimes a TAR(R) when concealing the dependence on k and d.

We proceed to estimate all our models using Conditional Least Squares Estimators (CLSE).[18] This is performed in a two-step process. Given values for the thresholds, each regime is estimated using OLS. Subsequently, we mininize the sum of the residual sum of squares over the values of γ_r using a simple grid-search.[19]

Asymptotic results are dependent on the continuity of the TAR model. In general, in an unrestricted TAR model, the conditional expectation of the dependent variable is a discontinuous function of the threshold variable at the threshold value. However, it can be easily verified that in equation (7) the restrictions guarantee that this function is continuous everywhere.

For the continuous TAR model the results of Chan and Tsay (1998) apply.[20] They show that under regularity conditions the CLSE of all parameters—*including* the threshold parameters—are $\sqrt{T}$-consistent and asymptotically normally distributed, and this permits the convenient construction of standard errors.[21] Although not explicitly shown, these results are assumed in the literature to generalize to threshold models with more than two regimes.

The results of Chan and Tsay (1998) are not widely recognized but they are important to the literature of market integration. In general one would expect that the conditional expectation function is continuous, but that the adjustment process is not. The fact that all inference is then standard makes life much simpler for the applied researcher. In contrast, inference on the thresholds in the general TAR model is currently only possible under further assumptions that we are unwilling to make.

[18]See, for instance, Chan (1993). Alternatively, one might estimate the models using Maximum Likelihood Estimation (MLE) assuming normality as in Qian (1998). The CLSE has attracted more theoretical attention and consequently more results are known. Moreover, CLSE identifies the thresholds from the change in slope parameters, while MLE also uses information in the change in variance. Since our model does not necessarily predict that the process v_t has constant variance, we prefer CLSE.

[19]Of course, the thresholds can not be more precisely identified than is allowed by the coarseness of the grid that is generated by the discrete (rounded) selection variable x_{t-d}. In our case, the exchange rate for the full sample has 234 unique values.

[20]Note that these results to hold, the continuity restrictions must be imposed on the estimated model.

[21]For the discontinuous TAR model the results of Chan (1993) apply. He shows in a two regime model that: (a) all the estimates of the unknown parameters are strongly consistent; (b) the thresholds converge at rate T; (c) the slopes parameters β_i^r and intercepts β_0^r are consistent at rate $\sqrt{T}$; (d) the slope parameters and thresholds estimates are asymptotically independent; and (e) the slope parameters are asymptotically distributed as if the thresholds are known.

4.1 A Fully Restricted Model

We first estimated the fully restricted model as in equation (7) using the full sample of the time series data, all trading days from 1879 to 1913 inclusive.[22] We restricted the grid search to values so that the middle regime by itself and the upper and lower regime combined have at least 5% of the observations. This corresponds to a restriction that the gold point falls in the interval [0.001,0.033] which seems a priori not unreasonable.

The results are presented in Table 3. The threshold is estimated to be $\gamma = 0.0324$ or 0.67 percent parity of parity, which is remarkably consistent with Officer's gold point estimates. The standard error on γ (computed using the Chan and Tsay (1998) methodology) is only 0.0013 which suggests that γ is very precisely estimated. Figure 3 plots the sum of squared residuals as a function of the threshold. The figure shows the surprising result that there exists a second local minimum at $\gamma = 0.010$ which is almost as low as the global minimum at $\gamma = 0.0324$.

4.2 An Unrestricted Model

To test the validity of the model, we then estimated a much less restricted specification. Specifically, without imposing any restrictions we estimated a series of TAR(R; k, d) models with $R = 3$, $1 \leq k \leq 9$; and $1 \leq d \leq k$. We then examined the following implications of the model: only two thresholds; symmetry of thresholds; convergence in outer regime towards the thresholds; symmetry in speed of adjustment in the two outer regimes; random walk behavior in the middle regime; and, finally, only a single lag in all regimes. To be consistent with the previous section, we imposed that the upper and lower regime combined have at least 5% and at most 95% of the observations.[23]

To choose between TAR models with the same R, but different values of k and d, various extensions of standard model-selection criteria have been proposed. We elected to minimize the BIC, defined as

$$\mathrm{BIC} = \sum_{r=1}^{R} n_r \ln \hat{\sigma}_r^2 + (k+1) \ln n_r$$

[22] In all the results that follow the first 10 observations are used as startup values and are omitted from the sample.

[23] This means that there is no restriction on the number of observations in each regime. If the parameters of either the lower or upper regime are unidentified, the residual sum of squares in that regime is set to zero.

Table 3: Restricted TAR(3) Constant Threshold Model

T		10616
$\log L$		50820.2
γ		0.032435
		(0.001378)
λ		-0.1069
		(0.0180)
Regime	T	SEE
Lower	81	0.00213
Middle	9949	0.00200
Upper	586	0.00213

Notes: See text. Standard error in parentheses. The equation estimated in each regime is equation 7.

where n_r is the number of observations falling in regime r, and $\hat{\sigma}_r^2 = \frac{1}{n_r} \sum_t (\hat{\epsilon}_t^r)^2$. This choice is motivated by the work of Kapetanios (1999) who showed that the standard results regarding the consistency of the BIC and Hannan-Quinn criteria and the non-consistency of the AIC procedure can been extended to TAR models. However, Wong and Li (1995) conjecture that the asymptotic efficiency of AIC in linear settings is carried over to TAR models, and simulation evidence seems to show that the AIC often performs better than BIC in small samples and is less likely to underestimate the lag length. Given the large sample size in our application, we suspect that the BIC will outperform other non-consistent criteria. We note that the BIC criterion generally biases the results towards more parsimonious models.

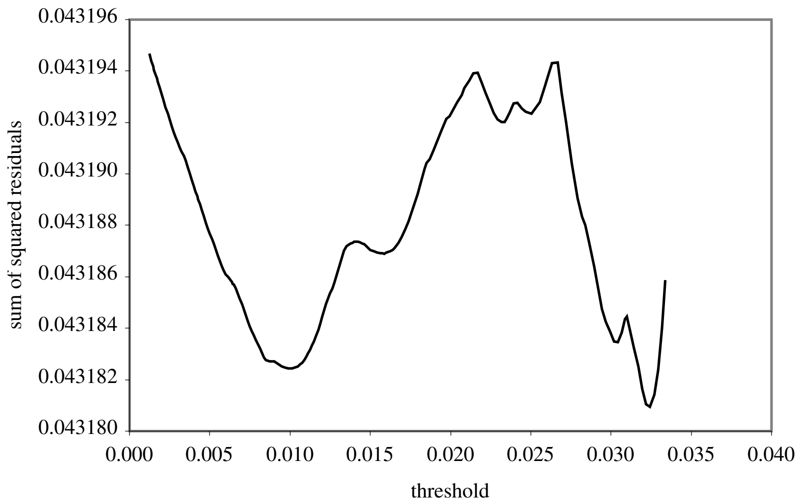
The BIC criterion selects the TAR(3;7,6) and we report the results in Table 4. The two thresholds are -0.0224 and 0.0003, respectively 0.46 percent below and 0.006 percent above parity. While the lower gold point estimate might seem reasonable the upper gold point estimate is too low to be credible. The upper regime also shows barely any convergence; in fact, it has a root closer to unity than the middle regime. Only the equality of the steady states and the threshold estimates seems to have some validity.

Under the assumption of a random walk in the middle regime, a test for the equality of the steady states of the outer regimes and the threshold estimates is simply a test of the continuous versus the discontinuous threshold model. We are unaware of any formal test for this purpose.

Testing for additional nonlinearities is not straightforward and has attracted very little attention in the literature.[24] The difficulty arises because the nuisance

[24] The test in Hansen (1999) excludes regime-dependent heteroskedasticity and can not be applied here.

Figure 3: Restricted TAR(3): Sum of Squared Residuals as a Function of the Threshold



Notes: See text and Table 3. The sum of squares shown is minimized over all other parameters.

parameters γ_r are not identified under the null, which makes the distribution of a likelihood ratio test nonstandard. This problem is sometimes referred to as the Davies problem (see Davies, 1977 and 1987). Two tests for testing linear versus nonlinear models that appear in the literature are those proposed by Luukkonen, Saikkonen, and Terasvirta (1988, henceforth LST), and Tsay (1989). LST construct tests of linear models versus Smooth Transition Autoregressive (STAR) models and show that these tests have power against TAR models; we use their Augmented First Order Test procedure: first estimate an AR(k) model, then regress the residuals on a limited set of second and third order terms, and test if this regression is statistically significant. Tsay's test exploits the similarity between change-point models and TAR models: after reordering the data based on the selection variable x_{t-d}, one tests for the orthogonality of the one-step ahead prediction error relative to the right-hand side variables.

Assuming that the thresholds are known, we can test for TAR(R) against TAR($R + S$) models by applying the above linearity tests on each regime. In the discontinuous TAR model the threshold estimates are rate-T consistent while slope estimates are only rate-$\sqrt{T}$ consistent and thus we can treat the thresholds as known and proceed accordingly. But if the TAR is continuous this argument

Table 4: Unrestricted TAR(3;7,6) Constant Threshold Model

T		10616	
$\log L$		51359.4	
γ_1		-0.022435	
γ_2		0.000315	
Regime	Lower	Middle	Upper
T	418	3857	6341
SEE	0.0036	0.0021	0.0017
β_0	-0.000908	-0.000030	-0.000009
	(0.001007)	(0.000063)	(0.000040)
β_1	1.002	1.061	1.117
	(0.050)	(0.016)	(0.012)
β_2	0.024	-0.053	-0.048
	(0.079)	(0.023)	(0.018)
β_3	-0.123	-0.042	-0.010
	(0.084)	(0.023)	(0.018)
β_4	0.033	0.002	-0.037
	(0.079)	(0.024)	(0.018)
β_5	0.088	0.000	-0.014
	(0.080)	(0.024)	(0.017)
β_6	-0.391	0.033	-0.011
	(0.083)	(0.024)	(0.018)
β_7	0.319	-0.015	-0.001
	(0.053)	(0.016)	(0.012)
$\sum_{i=1}^{7} \beta_i$	0.952	0.987	0.996
Steady state x	-0.0191	-0.0022	-0.0020
Tsay test	37.31	76.42	46.96
	[0.000]	[0.000]	[0.000]
LST	150.43	226.53	115.61
	[0.000]	[0.000]	[0.000]

Notes: See text. Standard errors in parentheses. p value in brackets. Standard errors are calculated under the assumptions of the discontinuous threshold model.

does not hold and we are unable to construct valid tests of remaining nonlinearities. Nevertheless, in Table 4 we report the Tsay and LST statistics. Although the significance level is likely to be overstated, the evidence suggests that remaining nonlinearities are somewhat problematic for the model as applied here.

Concluding, almost every implication of the economic model—single lag in AR processes, linearity after allowing for two thresholds, symmetry of thresholds, and symmetry in speed of adjustment—is strongly rejected. Only the unit root in the middle regime, and possibly the outer-regime convergence towards the thresholds have some validity.

4.3 A Non-Parametric Model

The results of the previous section are discouraging. We think that the problems are caused by declining thresholds, that is, an increase over time in "market integration."

The first strong indication of a changing degree of integration comes from Figure 4. It presents the standard deviation of x_t around the mean and around zero in a moving window of 1,000 observations. Both graphs show a strong decline over the entire period although the standard deviation around the mean shows a gradual decline over the entire period, while the standard deviation around zero shows a rapid decline during the second half of the period.[25]

Since we have no priors on how market integration improved over time, it is natural to think of a non-parametric approach. Specifically, we take a moving window of 1500 observations (about 5 years) and estimate a constant threshold model in each window.[26] First, to regain some confidence in the model after the disastrous results of the previous section we estimate an unrestricted model and test the symmetry of the thresholds, the unit root behavior in the middle regime, the convergence towards the thresholds in the outer regimes, and the symmetry of the speed of adjustment in the two outer regimes. We imposed the restriction of a single lag to limit the number of models that we needed to estimate. The symmetry of thresholds is tested in the following way: we estimate a model with symmetric thresholds but examine whether thresholds that are identified mainly from lower regime data are different from thresholds identified mainly from upper regime data.

So we estimate the model

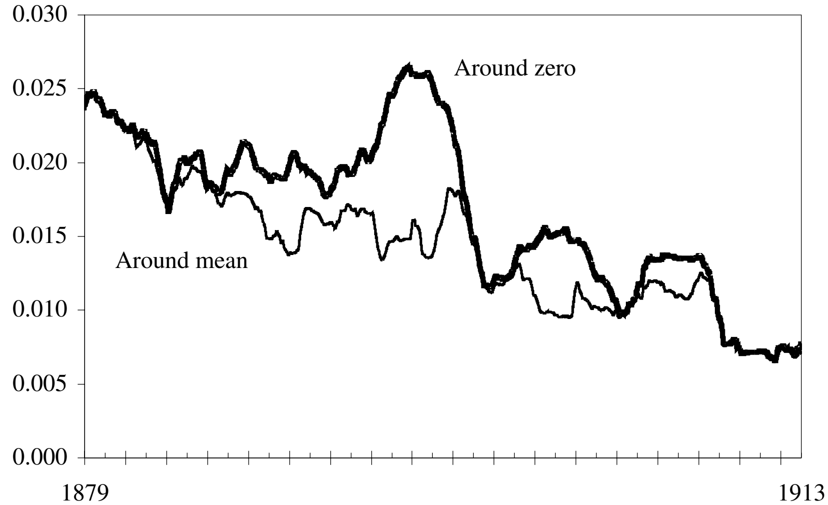
$$x_t = \begin{cases} \beta_0^1 + \beta_1^1 x_{t-i} + \epsilon_t^1, & x_{t-1} < -\gamma \\ \beta_0^2 + \beta_1^2 x_{t-i} + \epsilon_t^2, & -\gamma \leq x_{t-1} < \gamma \\ \beta_0^3 + \beta_1^3 x_{t-i} + \epsilon_t^3, & \gamma \leq x_{t-1} \end{cases} \tag{8}$$

As before we impose that the upper and lower thresholds combined have at least 5% and at most 95% of the observations. Since it is possible that during the 5 year period the exchange rate stayed within the middle regime most of the time we might not be able to identify the thresholds for each window. To purge thresholds

[25] These trends for daily data mimic the patterns seen in Officer's monthly data seen in Table 1.

[26] We took 1500 as a reasonable visual compromise between under- and oversmoothing. The window shifted by 10 observations each time to limit the computational costs. A single observation was eliminated because it was an extreme outlier and strongly affecting the results whenever it was present in the window.

Figure 4: Standard Deviation of Exchange Rate 1879–1913, 1,000 Day Moving Window



Source: *Financial Review*.

estimated from these cases we eliminate results when the LST test is not significant at the 10% level.

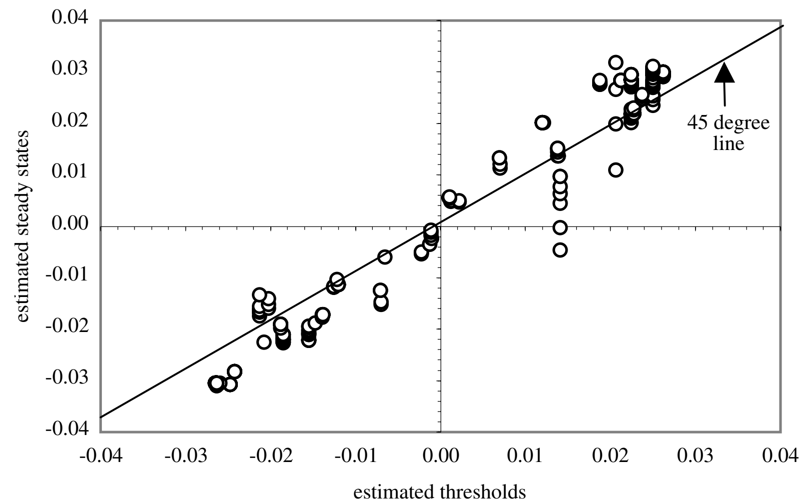
We start with an evaluation of the random walk prediction in the middle regime and the tests of convergence towards the thresholds in the outer regimes. Together, these predictions would imply that the threshold model is continuous. Since no formal tests exists for testing either prediction we proceed rather informally.

Concerning unit root behavior in the middle regime, we found the mean and standard deviation of β_1^2 were 0.9961 and 0.0420 respectively. Only 2.26% of the observations are higher than 1.025 or lower than 0.975. The estimated intercepts in the middle regime have a mean of -0.000012 and a standard deviation of 0.000057. Figure 5 shows the correspondence between directly estimated thresholds (γ) and the implied steady state of the outer regime in each sample.[27] The question is whether the outer-regime dynamics converge to the edge of the band, as in the continuous case. Looking across all of the sample windows, this property appears to hold.

We view these results as offering support for the continuous threshold model.

[27]Note that 6 of 483 points are not shown since they fall outside the plot area. These points show very different steady states from threshold estimates.

Figure 5: TAR(3) Model: Non-Parametric Thresholds and Steady States



Notes and Sources: See text.

We also compared the speed of adjustment in the two outer regimes. Here the model is not consistent with the data: the mean convergence rates are about 0.76 for the lower regime and 0.94 for the upper regime.

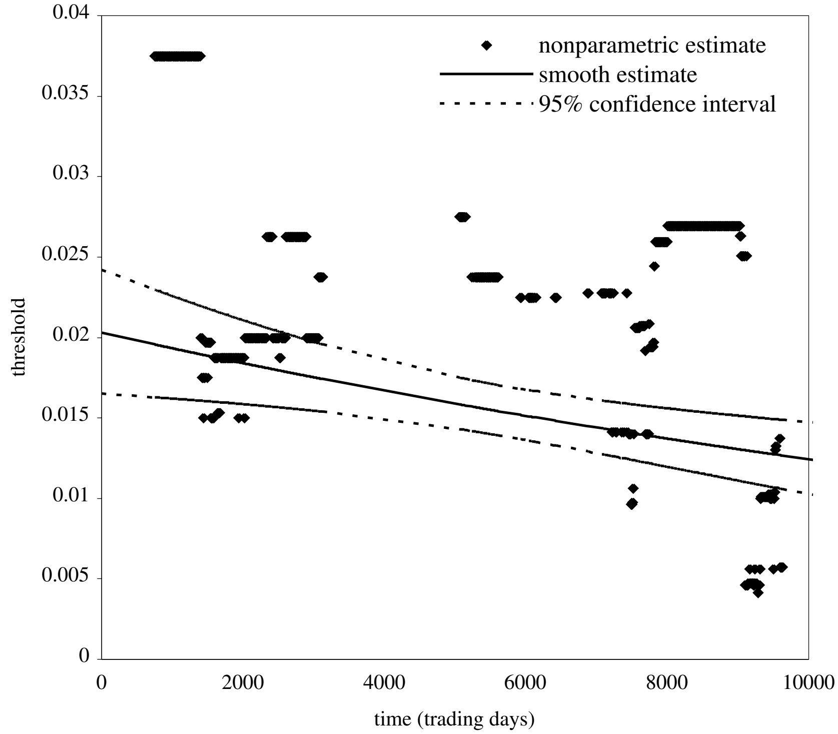
Based on these results we subsequently estimated nonparametrically a continuous threshold model similar to equation (7) relaxing only the equality of the λ parameters in the two outer regimes. The estimated thresholds are shown in Figure 6, and are labeled "nonparametric estimate."[28] Figure 6 provides evidence that the gold points experienced a dramatic decline during the period. Threshold estimates declined from 0.0375 at the start of the period to less than 0.010 in the end.

4.4 A Smooth Time-Trend Model

Given these findings, we prefer to model the dynamics of the gold points as a parametrically specified smooth time-trend during the entire period. Based on the results in the previous section, we estimate a fully restricted model as in equation (7) with the exception that we relax the symmetry of the speed of adjustment.

[28]The graph is very similar to a graph constructed from the threshold estimates in the unrestricted model.

Figure 6: TAR(3) Model: Nonparametric and Smooth Time-Trend Estimates



Notes and Sources: See text and Table 5.

$$\Delta x_t = \begin{cases} -\lambda_1(x_{t-1} - \gamma(t)) + \epsilon_t^1 & \text{when } x_{t-1} > \gamma(t) \\ \epsilon_t^2 & \text{when } \gamma(t) \geq x_{t-1} > -\gamma(t) \\ -\lambda_3(x_{t-1} + \gamma(t)) + \epsilon_t^3 & \text{when } -\gamma(t) > x_{t-1} \end{cases} \tag{9}$$

Again estimating this model using CLSE as explained earlier implies that $\gamma(t)$ has to be chosen using a grid search. This restricts the feasible number of unknown parameters in $\gamma(t)$. Although maybe not completely satisfactory considering Figure 6, we estimated such a model with only two unknown parameters: $\gamma(t) = \gamma_0 e^{-\gamma_1 t}$ with $t = 1, ..., 10615$.[29]

[29]The model remains a continuous model, and so we again rely on the Chan and Tsay (1998) methodology, adapted to this case. We started with a grid of 100 values for γ_0 between 0.00 and 0.05 and 100 values for γ_1 between -0.001 and 0.001. We then repeated the search, each time choosing an increasingly refined grid around the optimal point from the previous stage.

Table 5: TAR(3) Model where Thresholds Follow a Smooth Time Trend

T		10615	
$\log L$		51299.0	
γ_0		0.02036	
		(0.00197)	
γ_1		-0.00004879	
		(0.00001572)	
Regime	Lower	Middle	Upper
T	1144	6365	3106
SEE	0.002797713	0.002022852	0.001522273
Tsay test	6.810	0.297	5.555
	(0.033)	(0.862)	(0.062)
LST test	13.198	0.042	6.308
	(0.001)	(0.979)	(0.043)
λ	-0.0541		-0.0109
	0.0097		0.0026

Notes: See text. Standard errors in parentheses. p values in brackets. The threshold is modeled as $\gamma(t) = \gamma_0 e^{-\gamma_1 t}$.

The results are reported in Table 5. We also compute the Tsay and LST statistics for illustration. Strictly speaking these tests are invalid for the continuous model, and may overstate the significance level as before, but compared to the earlier results these statistics are now much lower, suggesting that there is little remaining nonlinearity.

That is, there seems to be some nonlinearity left, but this is unsurprising with more than 10,000 observations and the significance levels have decreased enormously from Table 4. The estimate of $\gamma(t)$ is plotted in Figure 6, labeled "smooth estimate." Perhaps surprisingly, this time trend is fairly flat compared to the non-parametric estimates. The time trend shows a decline of gold points from 0.0204 to 0.0121 during the full period while the non-parametric estimates decrease from 0.0375 to 0.0058. The non-parametric estimates suggest a very rapid decline in threshold size during the first decade (the 1880s) which is obviously not picked up by the smooth time-trend model. Either way, the evidence strongly points towards gold points that have significantly declined during the full period.

5 Gold Flows and Gold Point Estimates

Should we have much faith in this very different method for extracting adjustment dynamics, a method of revealed preference that derives implicit thresholds for quantity movements solely from the price dynamics? It would be nice to perform Newcomb's cross-check to see if our thresholds do a better job of predicting gold flows, so as to validate our model with flow data. Still, we are almost stymied in this regard since, for the most part, the extant flow data remain under a dark cloud of suspicion, as already noted (Morgenstern 1955; Goodhart 1969; Officer 1996).

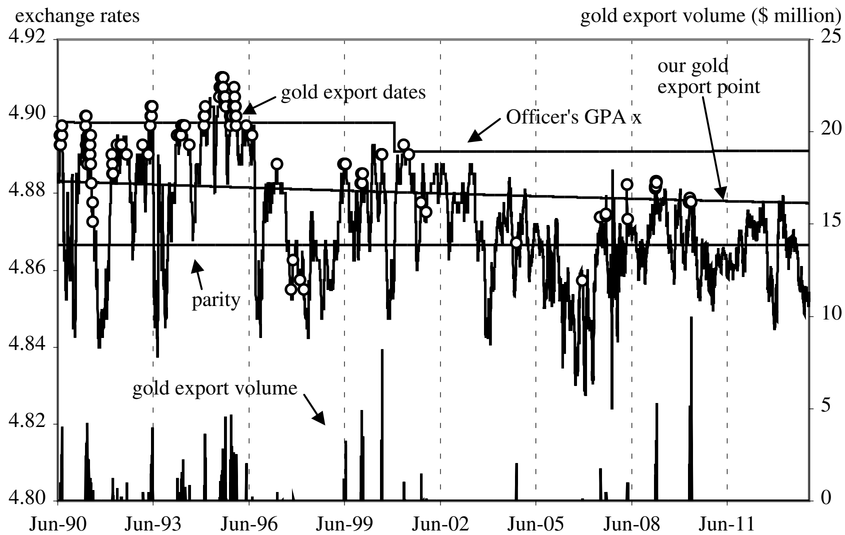
Nonetheless, we can now present a limited quantity-based cross-check on our method. In Figure 7, we plot for the time period for which we have Mint Report gold export data, the exchange rate, our gold export point estimates from the smooth time trend model, and Officer's export GPA gold point. With circles we indicate on the exchange rate line those days when substantial gold exports (more than $50,000) were observed. To account for the short lag in buying bills and shipping gold, we plot the maximum of the exchange rate in the three days before the actual shipment. On the bottom of the graph we also show the actual volume of the exports.

As mentioned before gold flow data are not very reliable. This is a possible reason for the five observations where gold was shipped to the U.K. when the exchange rate was actually below parity. It might also be the case that for these observations gold was shipped for other reasons than gold point arbitrage, a possibility that always renders problematic cross-checks of this kind. This explaination seems to be favored by the fact that the volumes shipped were small compared to other gold shipments. In any case, any single gold export observation should be viewed with skepticism and only general patterns can be considered relevant.

Our gold point estimates predict actual gold flows quite well for the period 1890-1896. For this period we observe almost no gold movement when exchange rates are within the bound and observe almost everytime some gold flow when the exchange rate is above the gold point. Arguably, however, the gold point might be slightly higher than our estimate (our non-parametric estimate is a little bit higher than the smooth trend estimate, and possibly more accurate). Officer's gold point, however, is almost certainly too high, except possibly for a short period in 1895. Around the turn of the century, our gold point estimate seems too low while Officer's estimate seems again too high. The truth appears to be somewhere in the middle. At the end of the period, between 1907 and 1910, it appears that we have hit the gold export point almost exactly right, although there are a couple of peaks that show no gold flows. Overall, by volume, only 35% of gold exports in this period occured at exchange rates above Officer's gold point; 50% of flows occurred above our estimate but below Officer's; and only 15% of flows occured below our estimate (with less than 1% below parity).

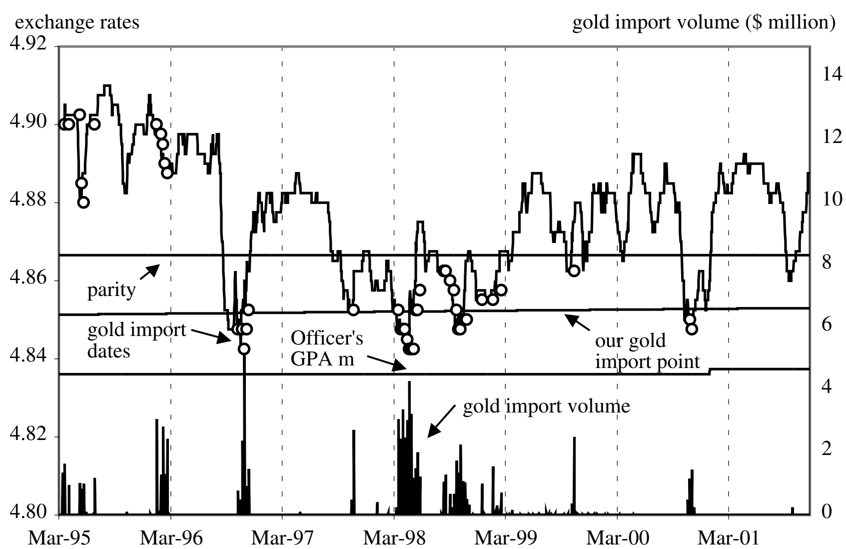
Figure 8 shows the same graph for Coleman's gold import data and the estimated

Figure 7: The Exchange Rate, Gold Points, and Gold Exports, 1890–1913



Notes and Sources: See text.

Figure 8: The Exchange Rate, Gold Points, and Gold Imports, 1895–1902



Notes and Sources: See text.

gold import points.[30] There are several reasons why this graph is less reliable than Figure 7. First, Coleman only collected weekly aggregates of gold imports. So the timing, of the gold imports is less precise and for this reason we have plotted the maximum exchange rate over ten days. Second, one has to account for the lag with which gold imports arrive. This lag is likely to be somewhere from one to two weeks. We have taken it to be ten days and so plotted with circles the maximum exchange rate from 19 days until 10 days before the end of the reporting week which showed gold imports. Third, Coleman collected the data from newspaper reports which are considered the least reliable source for gold shipments. These import data reveal some oddities. In 1895 and 1896 we see that large imports occurred at an exchange rate far above parity, but Coleman explains these as being driven by subscriptions to a large gold bond issue by the U.S Treasury, and hence such flows were unrelated to gold point arbitrage. We may also question the cause of the gold imports observed in late 1898 and early 1899 at exchange rates only just below 4.86. If these data are correct then our gold point estimates seem too high (that is, too far from parity). However, if these observations are also due to special circumstances then our estimates seem too be hitting the gold import point almost exactly. It is also clear that Officer's GPA gold import point is almost certainly too high. Overall, by volume, none of the gold imports in this period occured at exchange rates below Officer's gold point (indeed the exchange rate never fell so low); 47% of imports occurred below our estimate but above Officer's; and 52% of flows occured above our estimate (with 22% above parity, although most likely this large fraction of the flows was related to the gold bond issue and should be set to one side).

6 Conclusions

Our study offers several new issues for consideration. We have shown that work on the classical gold standard need not be confined to annual, quarterly, monthly, or even weekly data. Many financial publications list daily data, and we sample just one to construct a new series of dollar-sterling exchange rates. Such data are essential for the proper study of exchange rate dynamics where arbitrage operations were measured in days.

We then modeled the actual arbitrage process as described by the actors and experts at the time. We argued that increasing marginal costs of arbitrage are

[30]Only imports of more than $350,000 in a week are considered to be substantial, as indicated by circles.

essential to understanding the functioning of the system, in particular to explain the fact that persistent gold point "violations" and gold flows could be observed. Our model implies nonlinear dynamics and we used threshold autoregression methods to identify 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.

We compared our estimated thresholds with accepted estimates of the gold points, and found very different trends in the two measures. Our threshold declined dramatically but the measured gold points were fairly stable. We take this as evidence that the classical gold standard was an evolving standard in ways not very well captured by existing 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 convertibility commitment and the evolution of new technologies (or "learning") in the market itself. There might also have been increasing competition in the business of arbitrage itself, leading to decreases in the implied cost parameters.

There is some anecdotal evidence of changing costs over time, sometimes in ways dramatic enough to surprise even seasoned market participants. For example, there is an indication that, late in the period, banks trading on their own account were able to execute arbitrage, even in coin, at very low margins, perhaps below even their posted rates in the reports used today to construct direct gold point estimates. Thus, on May 20, 1909, the *New York Post* noted unusually heavy exports of gold coin, about $2.5 million, and observed that news of such shipments "was received with considerable surprise in financial circles, as it was not thought heavy exports of the metal could be made at prevailing exchange rates. The bankers making the shipments announced, however, that they were not losing money on the operation."[31] So much for received wisdom. One is relieved, of course, to learn that bankers had figured out how not to make a loss on simple arbitrage. At face value, this report offers strong support for our revealed-preference measures of the gold points—at the very least compared to method (a), consulting an expert!

Is such a report plausible? For reference, on May 19, the exchange rate stood at 4.87775, about 0.2 cents above our export point (4.8756) but still 2 cents below Officer's GPA export point (4.8909). Two tenths of a cent might not sound like much of a profit, but in this business the margins had certainly been shrinking, perhaps as a result of competition among banks. The correspondent for *Bankers Magazine*, could report on June 1910 (p. 924) that "there is not much profit in ship-

[31] Quotes here and below are as cited at <http://www.rms-republic.com>.

ping gold as might be thought, a thousand dollars on each million being considered quite enough of an inducement to make banking houses go in for transactions of this kind." Given the dollar-sterling parity, a margin of about 0.486 cents was apparently thought sufficient; by that reckoning, an extra 0.2 cents, as in May 1909, could be considered a significant extra incentive.

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 arithmetic 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, and opportunities abound to extend this type of study to other currencies or to markets for other goods.

References

Bai, J., "Estimating Multiple Breaks One at a Time." *Econometric Theory* 13 (1997), 315–52.

Bloomfield, A. I., *Monetary Policy Under the International Gold Standard: 1880–1914*. New York: Federal Reserve Bank of New York, 1959.

Chan, K.-S., "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model." *Annals of Statistics* 21 (1993), 520–33.

Clark, T. A., "Violations of the Gold Points, 1890–1908." *Journal of Political Economy* 92 (1984), 791–823.

Cole, A. H., "Evolution of the Foreign-Exchange Market of the United States." *Journal of Economic and Business History* 1 (1929), 348–421.

Coleman, A. M. G., "Arbitrage, Storage, and the 'Law of One Price.'" Princeton, N.J., March 1998. Photocopy.

Coleman, A. M. G., "Gold Flows and Triangular Arbitrage: Estimating the Dollar-Sterling Gold Points, 1886–1905." University of Michigan, October 2002. Photocopy.

Davies, R. B., "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative." *Biometrika* 64 (1977), 247–54.

Davies, R. B., "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternatives." *Biometrika* 74 (1987), 33–43.

Davis, L. E., and R. E. Gallman, *Evolving Financial Markets and International Capital Flows: Britain, The Americas, and Australia, 1865–1914*. Cambridge: Cambridge University Press, 2001.

Davis, L. E., and J. R. T. Hughes., "A Dollar-Sterling Exchange, 1803–1895." *Economic History Review* 13 (1960), 52–78.

Eichengreen, B. J., "Trends and Cycles in Foreign Lending." In *Capital Flows in the World Economy*, edited by Horst Siebert. Tübingen: J. C. B. Mohr, 1991.

Einzig, P., *The History of Foreign Exchange*. 2d 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.

Hansen, B. E., "Testing for Linearity." *Journal of Economic Surveys* 13 (1999) 551–76.

Kapetanios, G., "Model Selection in Threshold Models." *Journal of Time Series Analysis* 22 (2001), 733–54.

Krugman, P. R., "Target Zones and Exchange Rate Dynamics." *Quarterly Journal of Economics* 106 (1991), 669–82.

Luukkonen, R., P. Saikkonen, T. Teräsvirta., "Testing Linearity Against Smooth Transition Autoregressive Models." *Biometrika* 75 (1988), 491–99.

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, and A. M. Taylor, *Global Capital Markets: Integration, Crisis, and Growth*. Cambridge: Cambridge University Press, 2004.

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

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

Officer, L. H., "The Efficiency of the Dollar-Sterling Gold Standard, 1890–1908." *Journal of Political Economy* 94 (1986), 1038–73.

Officer, L. H., "The Remarkable Efficiency of the Dollar-Sterling Gold Standard, 1890–1906." *Journal of Economic History* 49 (1989), 1–41.

Officer, L. H., “Gold-Point Arbitrage and Uncovered Interest Arbitrage Under the 1925–1931 Dollar-Sterling Gold Standard.” *Explorations in Economic History* 30 (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.

Potter, S. M., “Nonlinear Time Series Modelling: An Introduction.” *Journal of Economic Surveys* 13 (1999), 505–28.

Prakash, G., “Pace of Market Integration.” Ph.D. dissertation, Northwestern University, June 1997.

Prakash, G, and A. M. Taylor., “Measuring Market Integration: A Model of Arbitrage with an Econometric Application to the Gold Standard, 1880–1913” Working Paper Series no. 6073, National Bureau of Economic Research, June 1997.

Qian, L., “On Maximum Likelihood Estimators for a Threshold Autoregression.” *Journal of Statistical Planning and Inference* 75 (1998), 21–46.

Seyd, E., *Bullion and Foreign Exchanges Theoretically and Practically Considered; Followed by a Defence of the Double Valuation, with Special Reference to the Proposed System of Universal Coinage*. London: Effingham Wilson, 1868.

Spalding, W. F., *Foreign Exchange and Foreign Bills in Theory and in Practice*. 2d ed. London: Sir Isaac Pitman & Sons Ltd., 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 (1988), 882–92.

Tong, H., *Threshold Models in Non-linear Time Series Analysis*. Berlin: Springer, 1983.

Tong, H.., *Non-Linear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press, 1990.

Tsay, R. S., “Testing and Modeling Threshold Autoregressive Processes.” *Journal of the American Statistical Association* 84 (1989), 231–40.

Wong, C. S., and W. K. Li, “A Note on the Corrected Akaike Information Criterion for Threshold Autoregressive Models.” *Journal of Time Series Analysis* 19 (1998), 113–24.