Measuring Market Integration: Foreign Exchange Arbitrage and The Gold Standard, 1879–1913

Eugene Canjels
New School University[*]

Gauri Prakash-Canjels
Howrey and Simon[†]

Alan M. Taylor
University of California at Davis and NBER[‡]

May 2002

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 nonlinear theoretical models applied using the techniques of threshold autoregressions. We 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 implied gold points and price dynamics. The changes in these measures over time provides an insight into the evolution of market integration.

---

[*]Assistant Professor, Department of Economics, Graduate Faculty, New School University, 80 Fifth Avenue, 5th floor, New York, NY 10011. Tel: (212) 229-5901. Fax: (212) 229-5903. Email: <canjelse@newschool.edu>.

[†]Senior Economic Manager, Capital Affiliates, Howrey, Simon, Arnold & White, 1299 Pennsylvania Avenue NW, Washington, DC 20004. Tel: 202-383-6815. Fax: 202-383-7237. Email: <prakash_canjels@yahoo.com>.

[‡]Associate Professor, Department of Economics, University of California, One Shields Avenue, Davis, CA 95616-8578. Tel: 530-752-1572. Fax: 530-752-9382. Email: <amtaylor@ucdavis.edu>. WWW: <http://www.econ.ucdavis.edu/faculty/amtaylor/>.

1 Introduction

How can we measure market integration? By way of illustrating a new approach to the problem, this paper is concerned with assessing 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 be expected to 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. Yet establishing such a connection remains vital in any assessment of market integration, in this case just as in any other. The low frequency of the extant exchange rate data and the low quality of the extant gold flow data impeded previous researchers in this regard. To try to settle the issue we collected new data for the highest frequency and the longest duration possible for this market—daily prices for the entire period of dollar-sterling convertibility from 1879 to 1913.

By analyzing high-frequency price dynamics using nonlinear dynamic models, we seek to establish that the evolution of gold point arbitrage was rather different than has been understood up to now. To corroborate this revisionist interpretation, we find support by drawing on previously unused data on gold flows for certain subperiods. We find strong evidence in price behavior of a gradual evolution towards increased market integration as time wore on, in contrast to the traditional view in which the gold points are viewed as having been more or less stable over time.

1.1 Gold Point Controversies

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]

[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).

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 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 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, 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 1998). The conventional view also accords with contemporary observations from the period, wherein attention given to the rise in foreign exchange operations became so marked that the term "arbitrage" came into common usage.[2]

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 two ways. First, we can try to better formalize the theory of arbitrage in 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.

1.2 The State of the Art

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.[3]

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 available. What are those alternative methods? Officer's nine categories can be collapsed into four:

> **(a) Consult an Expert or Two.** A century ago the most widely-used estimates of the gold points published in newspapers relied on these less scientific, informal methods. An "expert" typically meant an experienced market participant such as a foreign-exchange dealer or a banker. But as Officer notes, such sources may be inconsistent over time, unrepresentative, omit some cost components, 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 as a means to assess the workings of the exchange by Newcomb (1886, pp. 281–82), but his suggestions were neglected at the time. The method has recently been applied by various authors (e.g., Clark 1984), but has run into practical difficulties due to the poor quality of gold-flow

[2] For example, the Oxford English Dictionary (Second Edition, 1989, <http://dictionary.oed.com/>) cites the first instance of the word *arbitrage* in its economic sense as occurring in the 2nd *Encyclopædia Britannica* of 1875, with reference to British government securities. Only a few years later the *Daily News* of 1881 is recorded to have applied the term to the market for Bills of Exchange, and it is arbitrage in these financial instruments that the OED chooses to emphasize as the principal usage of the word: "arbitrage, *n.* 3. *Comm.* The traffic in Bills of Exchange drawn on sundry places, and bought and sold in sight of the daily quotations of rates in the several markets, each operation being based in theory on the calculation known as Arbitration of Exchange, q.v."

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

data. 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. 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: we will show evidence of persistent disequlibria occurring.[4]

(d) Break Down All Individual Components of Transaction Costs. Officer prefers this method. The benefit is that it makes no maintained theoretical assumptions. The obvious problem is the intense data requirements. A less-obvious problem is whether these direct cost estimates do indeed correspond to the behavior of the arbitrageurs in question.

It is a testament to Officer's determination and great scholarship that, having pointed out the drawbacks of methods (a), (b) and (c), he is able to show the feasibility of method (d) above. Future generations of scholars will benefit from the detailed cost estimates Officer has constructed in his studies of the exchange. Along the way, 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.[5]

Still, it remains an open question as to 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). This would allow us to 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. 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 in a way that detects the point at which we see the onset of arbitrage behavior, 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. For example, efforts to study price dynamics using method (c), have suffered from the use of monthly data. Such data are obviously inadequate even for that task, since the maxima and minima of monthly data may dramatically understate the

[4]Even if it were valid, there are other practical problems with this method as it has been employed to date. problem with this method 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.

[5]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.

volatility of prices in a market where shipping times were calculated in terms of days for the Atlantic crossing. 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

We first describe the new data on exchange rates we have compiled which will allow us to examine the workings of the classical gold standard at a much a higher frequency than any earlier work. Previous researchers have relied on monthly, quarterly, or annual data series. Even so assiduous a compiler of data as Officer (1996, chap. 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 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).

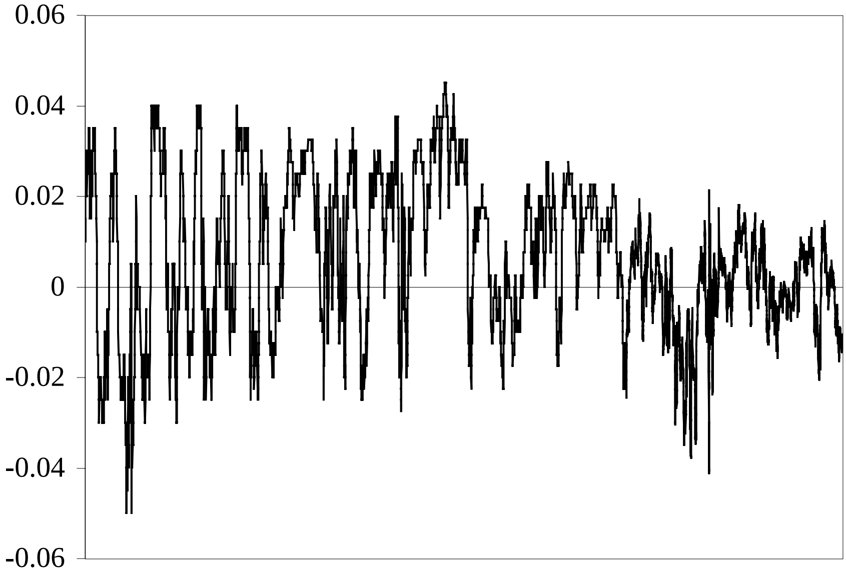
We collected the exchange rates based on sterling demand bills in New York since it is now accepted that the dominant form of arbitrage under the classical gold standard was via the demand bill (or sight bill) denominated in sterling and drawn on London. Only later, in the interwar period, did cable 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).[6]

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.[7] Sight

[6] 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. Recall 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 price that was certain. Thus forward cover was included in the bill, but it still was subject to discount vis-à-vis cable 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, pp. 60–63 and 115).

[7] 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

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



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

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.[8] 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 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 data set we should have ample information to identify parameters of interest and how they change over time.

Figure 1 displays the exchange rate deviation $Y_t = E_t - E^{\mathrm{par}}$ measured in dollars where we define the parity exchange rate 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),

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."

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

that is, $E^{\text{par}} = P^g/P^{g*}$.[9]

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 Y_t will be the object of study in the remainder of the paper. We think it is fairly clear that Y_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.[10]

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 Morgenstern (1955). Morgenstern 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 on this data 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 caused substantial headaches 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.[11] 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.[12] He writes: "I found the French

[9] For the period under consideration $E^{\text{par}} = 4.86656$.

[10] 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.

[11] These tables are labeled "Statement of United States Gold Coin and Gold Bullion Exported from the Port of New York to Europe during the Fiscal Year, [Year]".

[12] 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

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. 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. Obviously, if this hitherto neglected source might be reliable enough to launch such a major commercial endeavor, it behooves any scholar interested in the gold standard to place the data under some scrutiny too.

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 off 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

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). On top of all this Bayerle now finds an "apparent collaboration by five governments and eight financial institutions" to conceal the loss of the gold and efforts to hide the true location of the watery grave off Nantucket. Hence, since locating the wreck in 1981, he has exhaustively explored every single source of gold flow data to show that on a day-to-day basis he can match all the "engagements" of shipping lines to move gold out of the U.S. with the corresponding Assay Office and overseas flow data, with sufficient accuracy to prove that the gold was probably lost on the ill-fated voyage. The story about the ship and the analysis of the gold flow data can be found on the web-site <http://www.rms-republic.com>.

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.[13] 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.[14]

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. 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).[15]

[13] 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 also been a subject of considerable debate (Davis and Hughes 1960; Officer 1996).

[14] 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. 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 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.

[15] Note that we also neglect other centers besides New York and London. These were two very big centers, and to a first approximation this restriction may be reasonable. It also accords with the classical view of the gold standard as a "hub and spoke" system centered on London, where, it terms of arbitrage and the exchange rate adjustment of the whole system, the Bank of England stood, in Keynes' felicitous phrase, as the conductor of the international orchestra.

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 $£P^{g*}$
$t = 0$	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 $E < E^{par}$				
Time	Place	Action by arbitrageur	Gain	Loss
$t = 0$	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).

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 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. 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. 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.[16] 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 immediately realized.[17] 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.[18] 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.[19]

3.3 Costs of Arbitrage

The final building block 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.[20] 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

[16] In contrast, Coleman's (1998) model includes shipment delays as would be seen in typical commodity markets.

[17] 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 $t = 0$.

[18] 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 $2T$.

[19] To 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 that is tangential to gold-standard arbitrage as it is purely construed, and we omit it in the present analysis. At the margin we are focusing *only* on the arbitrage processes described in Table 2.

[20] As just shown, the quantities of bills and gold involved in arbitrage are proportional, so either might be used to measure the flow.

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 illustrate as depicted in Figure 2.

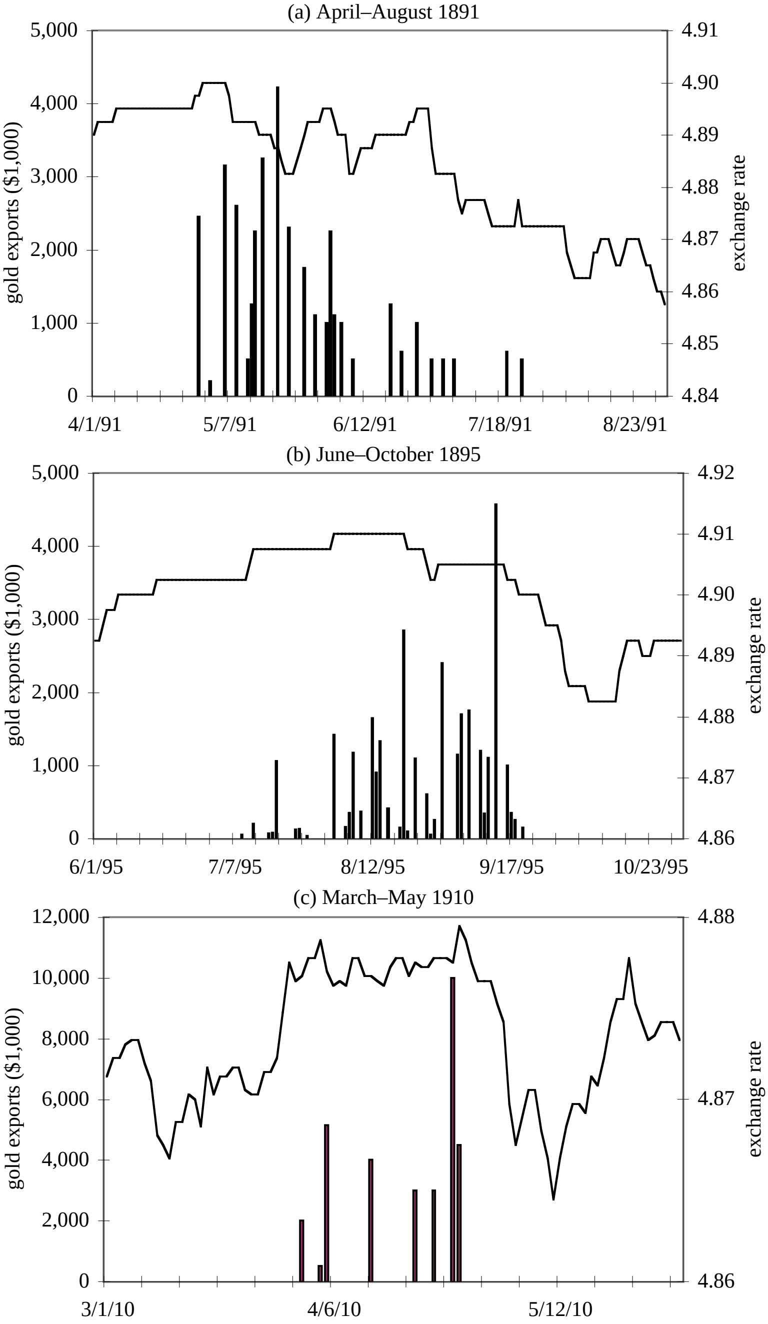
The first period of heavy gold exports was from April 29, 1891 to July 6, 1891; the second period was from July 9, 1895, to September 20, 1895; and the third period was from April 1, 1910, to April 26, 1910. In each of these episodes \$30 to \$35 million was exported from New York to England. As can 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 \$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 that one of the main reasons for these results is that the marginal cost of gold point arbitrage increased with the quantity shipped. This could have at least two causes.

Firstly, the arbitrage firms had a limited amount of capital available for their operations and using it for GPA makes it unavailable for other usage. If they have a portfolio choice of where to invest their capital they will first divert 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.[21] However, when exports where heavy the U.S. Mint could not always provide gold bars, and arbitrageurs would have

[21] 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).

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



Sources: *Commercial and Financial Chronicle* and U.S. Mint Reports.

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." On the same day *New York Commercial* reported that "this engagement of gold bars again exhausts the supply at the Assay Office and further exports will have to be made in gold coin." Such events were not unusual. Similar conditions had obtained a year earlier when the same newspaper had reported on December 23, 1908, that "if the demand for London remittances continues to increase, it is believed that the sterling exchange price will reach figures at which exports of gold coin will become profitable. The available supply of gold bars in the United States Assay Office now is exhausted…." And 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 forced to place limits on such tactics.[22]

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 deployment of such "gold devices" could have also thrown sand in the wheels of gold point arbitrage, leading just as surely to increasing marginal costs for individual arbitrageurs.

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.

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

3.4 Optimal Arbitrage

We now consider the arbitrage operation. Arbitrageurs will only move gold up to the point where marginal revenue equals marginal cost. Without loss of generality consider a case of 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 given by $MR = (E_t - E^{\text{par}})P^{g*}$; 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 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 $v_t \sim N(0, \sigma^2)$. However, our arguments will generalize to other processes for v_t.

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, namely

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

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, 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, 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. Otherwise, we are in a regime where adjustment takes place with a longer lag than one period. This would seem empirically relevant given the known lags in the actual arbitrage process.

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, 1999a, for an overview).

The general form of a TAR model can be expressed as

$$Y_t = \beta_0^r + \sum_{i=1}^{k} \beta_i^r Y_{t-i} + \epsilon_t^r, \tag{8}$$

where

$$\begin{aligned} r = 1 \quad &\text{if} \quad \gamma_0 < Y_{t-d} \leq \gamma_1; \\ r = 2 \quad &\text{if} \quad \gamma_1 < Y_{t-d} \leq \gamma_2; \\ &\ldots \\ r = R \quad &\text{if} \quad \gamma_{R-1} < Y_{t-d} \leq \gamma_R; \end{aligned}$$

and ϵ_t^r is a martingale difference sequence with a regime-dependent variance σ_r^2. In this model AR(k) dynamics obtain in regime r governed by parameters (β_i^r, σ_r^2). The process is in regime r at time t when the selection variable Y_{t-d} lies in the interval between two thresholds $\gamma_{r-1} < Y_{t-d} \leq \gamma_r$, 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 TAR(R) if the values of k and d are not relevant in the context that it is used. The parameter R is the number of regimes, the parameter k indicates the number of AR lags in each regime, and the delay parameter d is the number of lags in Y_t that determine the current regime position. Note that a TAR(1) is the same as an AR model.

For future reference we can use two alternative representations of the model. First, we may write, from the above

$$Y_t = X_t' \beta^r + \epsilon_t^r, \tag{9}$$

where $X_t' = (1, Y_{t-1}, ..., Y_{t-k})$. Second, it is sometimes useful to order the observations by the value of the selection variable, the lagged Y_{t-d}. This allows the model to be rewritten as an ordered regression

$$\tilde{Y}_s = \tilde{X}_s' \beta^r + \tilde{\epsilon}_s^r, \tag{10}$$

where the data vector $(\tilde{Y}_s, \tilde{X}_s)$ is simply the data vector (Y_t, X_t) ordered on Y_{t-d}. Why is the ordered regression useful? Suppose that

$$\begin{aligned}
-\infty = \gamma_0 < \quad & \tilde{Y}_1 \leq \ldots \leq \tilde{Y}_{s_1} & \leq \gamma_1; \\
\gamma_1 < \quad & \tilde{Y}_{s_1+1} \leq \ldots \leq \tilde{Y}_{s_2} & \leq \gamma_2; \\
\ldots & & \\
\gamma_{R-1} < \quad & \tilde{Y}_{s_{R-1}+1} \leq \ldots \leq \tilde{Y}_T & \leq \gamma_R = +\infty.
\end{aligned} \tag{11}$$

Then for $s = 1, \ldots, s_1$ all of the observations $(\tilde{Y}_s, \tilde{X}_s)$ lie in the regime $r = 1$; for $s = s_1 + 1, \ldots, s_2$ all of the observations lie in the regime $r = 2$, and so on. Estimation of the thresholds γ_r is thus equivalent to estimating the parameters s_r that split the samples in the ordered regression. This notation also clarifies the close correspondence between change-point and TAR models.

Note that we allow the variance of the error term to be regime dependent, a feature that is not common to all TAR models in the current literature. This is warranted by our theoretical model and, as a practical matter, we find that there is substantial difference in the variance of shocks across regimes in our case. However, these modeling benefits do come at a cost, since the assumption of regime-dependent heteroskedasticity imposes some restrictions on the statistical techniques that can be used later on.

The models are estimated using Maximum Likelihood Estimation (MLE) under the assumption that the errors have an i.i.d. normal distribution. This is most easily performed by first concentrating the likelihood function. Conditional on the thresholds γ_r, MLE is identical to OLS in each regime. We then maximize the likelihood by a grid search over all possible values of the thresholds γ_r.[23]

To perform estimation reliably, we arbitrarily set what we thought were reasonable restrictions on how many data points were need to identify parameters in a single regime by requiring at least 20 observations in each regime. As regards other aspects of model selection, we elected to use the Schwarz Criterion (SC) to select the lag length k and the delay parameter d. The Schwarz Criterion was purposely chosen because it leads to more parsimonious models than the Akaike Information Criterion.

4.1 Constant Gold Points?

The classical gold standard period is generally considered to be a period where the gold points were fairly stable. In light of this we first estimate models with constant gold points over time. The possibility of time-varying gold points will be discussed in the next section.

We first estimated a series of TAR($R; k, d$) models with $R = 3$, $1 \leq k \leq 6$; and $1 \leq d \leq k$. We applied these models to the full sample of the time series data, all trading days from 1879 to 1913 inclusive.[24] For that sample, the model that minimized SC was a TAR(3;5,5). The estimation results of this model are reported in Table 3. The two thresholds are -0.0202 and 0.0036, respectively 0.42 percent below and 0.07 percent above parity. While the lower gold point estimate might seem reasonable the upper gold point estimate is almost certainly too low and implausible. It is interesting that the for all the estimated TAR models the lower gold point estimate was very close to -0.02. The upper gold point estimate, however, was either 0.0203, 0.0136, or close to 0.0036.

Figure 3 shows the condensed likelihood function. That is, it plots the maximum likelihood as a function of the location of one of the thresholds, maximizing over the other threshold and the regression parameters. The function is wildly erratic and suggests the existence of additional nonlinearities. Testing for such additional nonlinearities is not straightforward and there are no appropriate tests available for testing TAR(R) against TAR(R+S) models.[25]

To develop ideas, we note that testing for linear versus TAR models is already nontrivial. The difficulty arises because the nuisance 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 methods for testing linear versus nonlinear models that appear in the literature are the Augmented First Order Test proposed by Luukkonen, Saikkonen, and Teräsvirta

[23] 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 Y_{t-d}. In our case, rounding in the data limits the grid to at most 233 discrete possibilities for TAR(2) models and at most $233 \times 232/2 = 27{,}028$ distinct possibilities for TAR(3) models. The $\hat{\gamma}_r$ that we report are the midpoints between $\tilde{Y}_{s_r}$ and $\tilde{Y}_{s_r+1}$. For similar approaches see Fanizza (1990) and Balke and Fomby (1997). For more details on the algorithm see the appendix of our earlier draft (Prakash and Taylor 1997).

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

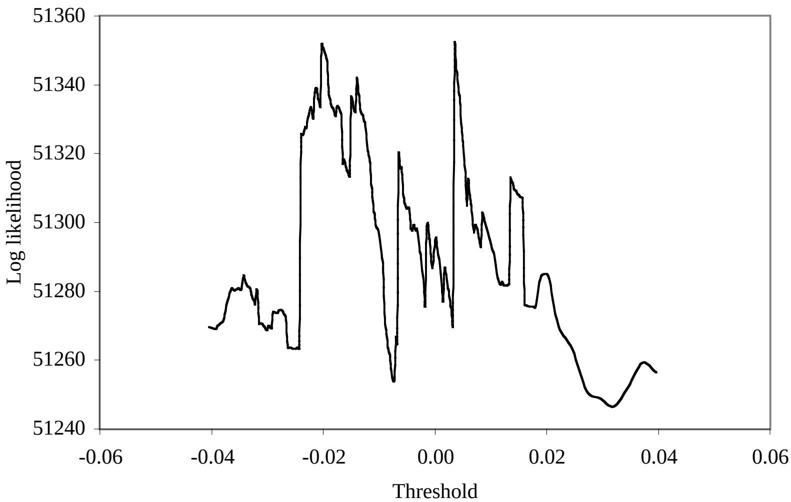
[25] The only test that the authors are aware of is in Hansen (1999), but since it excludes regime-dependent heteroskedasticity it cannot be implemented here.

Table 3: The Preferred TAR(3) Model

T		10616	
log *L*		51351.6	
SC		-9.6543	
Thresholds (γ)		-0.020185	
		0.003565	
Regime	Lower	Middle	Upper
T	722	4486	5408
SEE	0.0033	0.0021	0.0017
β_0	-0.000843	-0.000042	0.000011
	(0.0006)	(0.0000)	(0.0001)
β_1	1.005	1.101	1.094
	(0.038)	(0.015)	(0.013)
β_2	0.004	-0.066	-0.037
	(0.057)	(0.022)	(0.020)
β_3	-0.059	-0.042	-0.021
	(0.059)	(0.022)	(0.019)
β_4	0.015	-0.005	-0.021
	(0.058)	(0.023)	(0.018)
β_5	-0.017	-0.002	-0.021
	(0.043)	(0.016)	(0.013)

Notes: See text. The equation estimated in each regime is equation 8 in the text.

Figure 3: The TAR(3) Condensed Likelihood Function



Notes: See text and Table 3.

Table 4: Testing for Nonlinearity in the Preferred TAR(3) Model

	LST test		Tsay test	
	Test statistic	Probability	Test statistic	Probability
Lower	181.56	0.00	40.10	0.00
Middle	102.11	0.00	30.55	0.00
Upper	86.22	0.00	10.05	0.12

Notes: See text. LST is the Augmented First Order Test by Luukkonen, Saikkonen, and Teräsvirta (1988); Tsay test is the χ^2 test of Tsay (1989). For the Tsay test we use $b = T/10$ observations for the first iteration.

(1988, henceforth LST), and the test proposed by Tsay (1989).[26]

LST construct tests of linear models versus Smooth Transition Autoregressive (STAR) models. While this approach is unlikely to be optimal (that is, most powerful) because the alternative is not correctly specified, the test was shown by LST to have power against TAR models. Following LST we use the Augmented First Order Test procedure: first regress y_t on an intercept and k lags, to estimate the AR model. Then regress the residuals on an intercept, the k lags, the second order terms $y_{t-i}y_{t-j}$ for $i, j \leq k$, and a limited set of third order terms y_{t-i}^3 for $i \leq k$. The linearity assumption implies that the second and third order terms should have zero coefficients and this can be tested using a regular F or Wald test.

The test in Tsay (1989) exploits the similarity between change-point models and TAR models. The test is based on the ordered regression representation. Under the null the model exhibits parameter constancy, while under the alternative the parameters change with "time" (that is, re-ordered time). Hence, Tsay 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. That is, if the correct number of thresholds has been chosen, it will be just enough to purge nonlinearity from each regime. The difficulty arises from the fact that the thresholds are estimated, but given that the thresholds are rate-n consistent we expect this procedure to be reasonably satisfactory even when the thresholds are estimated, as here, rather than known.

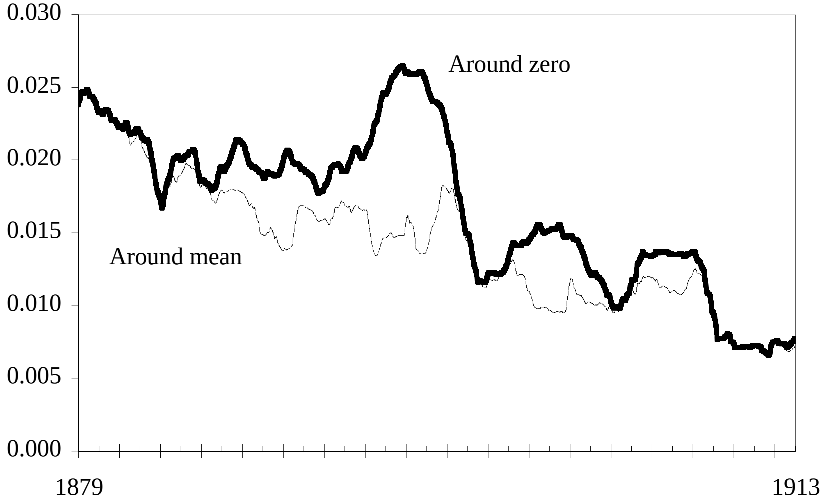
The results in Table 4 show that within each regime the linear model is easily rejected for all regimes using the LST test and for two out of three regimes using the Tsay test.

4.2 Evolving Gold Points

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

[26] We also considered implementing the testing procedures outlined by Hansen (1996; 1999). He considers regular Wald statistics and provides simulation methods to estimate the nonstandard distribution of these statistics. Unfortunately, these simulations are computationally expensive and infeasible for our application. For example, trying to reproduce Figure 3 in Hansen (1999) we found that one simulation with our data set of over 10,000 points took about 60 times as long as a simulation with the original sunspot data consisting of only 278 observations.

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



Source: *Commercial and Financial Chronicle*.

The first strong indication of a such a trend comes from Figure 4, which presents the standard deviation of Y_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 from 1879 to 1913 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.[27]

How can we model changing thresholds over time? The approaches taken in the change-point literature are not feasible here due to computational restrictions. We elected to split the data into a set of equal-sized subperiods and re-estimate the model. The number of intervals was allowed to vary between 1 (the constant parameter case discussed above) to 17 (chosen so that each interval represents at least two years and 600 observations). For each possible split we estimated the full range of AR and TAR models discussed earlier. The split that minimizes the Schwarz Criterion has 13 different subperiods and the model is, in most cases, the TAR(2;1,1). Table 5 presents the results.

It may seem surprising that we find here only one threshold in each period, that is, two regimes, a TAR(2). However, this is due to the small samples in each subperiod. The exchange rates typically did not vary enough to exhibit adjustments on both sides of the band in any of the subperiods. Equivalently, in most periods we find that the exchange rate crossed over the gold import point or the export point, but rarely both. Only rarely, then, is it possible to identify a second threshold, and thus estimate both gold points. Most often we are left able to model either the lower and middle regimes, or middle and upper regimes, but not all three.

[27] The trends in this figure for daily data exhibit the same trends seen in Officer's monthly data shown in Table 1.

Table 5: TAR Models in 13 Subperiods

Period	1	2	3	3 TAR(3)	4	5	6	7
Start	1.14.1879	9.16.1881	5.21.1884	5.21.1884	1.24.1887	9.27.1889	6.3.1892	2.12.1895
End	9.15.1881	5.20.1884	1.22.1887	1.22.1887	9.26.1889	6.2.1892	2.11.1895	10.18.1897
T	816	816	816	816	816	816	816	816
log *L*	3639.07	3786.84	3705.03	3712.85	4210.68	3814.27	3929.22	4213.99
SC	-8.862	-9.224	-9.023	-9.010	-10.263	-9.291	-9.573	-10.271
Threshold(s)	-0.039060	0.020940	0.029690	-0.012810	0.014690	0.014690	0.017190	0.012190
				0.029690				
Lower Regime								
T	25	—	—	189	—	—	—	—
SEE	0.00508	—	—	0.00248	—	—	—	—
β_0	-0.031555	—	—	-0.002428	—	—	—	—
	(0.01281)	—	—	(0.00103)	—	—	—	—
β_1	0.255	—	—	0.849	—	—	—	—
	(0.291)	—	—	(0.054)	—	—	—	—
Halflife	1	—	—	5	—	—	—	—
LST test	0.87*	—	—	5.69	—	—	—	—
(prob)	(0.351)	—	—	(0.058)	—	—	—	—
Tsay test	NA	—	—	5.91	—	—	—	—
(prob)	NA	—	—	(0.052)	—	—	—	—
Steady State	-0.0423	—	—	-0.0161	—	—	—	—
	(0.0015)	—	—	(0.0015)	—	—	—	—
Convergence t-test	-2.19	—	—	-2.14	—	—	—	—
Middle regime								
T	791	485	708	519	301	499	349	198
SEE	0.00275	0.00263	0.00281	0.00290	0.00193	0.00273	0.00283	0.00194
β_0	-0.000111	-0.000032	0.000059	-0.000073	0.000106	-0.000043	0.000191	-0.000100
	(0.00010)	(0.00014)	(0.00011)	(0.00014)	(0.00011)	(0.00013)	(0.00016)	(0.00015)
β_1	0.994	0.984	0.992	1.005	0.999	0.980	0.985	0.990
	(0.004)	(0.010)	(0.007)	(0.010)	(0.010)	(0.011)	(0.013)	(0.013)
Halflife	126	43	92	NA	711	34	46	67
LST test	5.21	0.89	9.40	1.76	5.21	0.38	1.97	3.07
(prob)	(0.074)	(0.640)	(0.009)	(0.416)	(0.074)	(0.826)	(0.373)	(0.216)
Tsay test	2.78	0.51	13.84	NA	3.22	0.45	0.52	2.35
(prob)	(0.248)	(0.774)	(0.001)	NA	(0.200)	(0.797)	(0.769)	(0.309)
Steady State	-0.0202	-0.0020	0.0078	0.0137	0.1091	-0.0021	0.0127	-0.0096
	(0.0221)	(0.0082)	(0.0165)	(0.0290)	(1.1479)	(0.0061)	(0.0127)	(0.0155)
Upper regime								
T	—	331	108	108	515	317	467	618
SEE	—	0.00197	0.00151	0.00151	0.00115	0.00169	0.00150	0.00124
β_0	—	0.001135	0.015713	0.015713	0.000212	0.000765	0.000287	0.000055
	—	(0.00059)	(0.00463)	(0.00463)	(0.00024)	(0.00047)	(0.00041)	(0.00016)
β_1	—	0.962	0.517	0.517	0.989	0.963	0.985	0.996
	—	(0.018)	(0.138)	(0.138)	(0.010)	(0.021)	(0.016)	(0.006)
Halflife	—	18	2	2	65	19	47	185
LST test	—	2.16	0.97*	0.97*	1.13	1.25	2.30	0.46
(prob)	—	(0.339)	(0.324)	(0.324)	(0.569)	(0.535)	(0.317)	(0.795)
Tsay test	—	NA	0.85	0.85	0.76	0.21	2.41	0.06
(prob)	—	NA	(0.653)	(0.653)	(0.685)	(0.899)	(0.299)	(0.972)
Steady State	—	0.0299	0.0326	0.0326	0.0198	0.0205	0.0195	0.0146
	—	(0.0030)	(0.0004)	(0.0004)	(0.0061)	(0.0027)	(0.0084)	(0.0226)
Convergence t-test	—	3.03	7.35	7.35	0.84	2.10	0.27	0.11

Notes: See text. The equation estimated in each regime is equation 8 in the text. * denotes simple first-order statistic. Overall: $\log L = 52753$, $SC = -9.86$.

Table 5 (continued) TAR Models in 13 Subperiods (continued)

Period	8	9	10	11	11 TAR(3)	12	13	13 TAR(3)
Start	10.19.1897	07.06.1900	03.19.1903	11.23.1905	11.23.1905	08.01.1908	04.14.1911	04.14.1911
End	07.05.1900	03.18.1903	11.22.1905	07.31.1908	07.31.1908	04.13.1911	12.31.1913	12.31.1913
T	816	816	816	816	816	816	824	824
log *L*	4138.75	4330.01	4299.62	3693.88	3770.59	4341.69	4450.74	4459.42
SC	-10.086	-10.555	-10.481	-8.996	-9.151	-10.584	-10.746	-10.734
Threshold(s)	-0.017810	0.007190	-0.002685	0.006565	-0.014935	0.009065	-0.006685	-0.005935
					0.006565			0.006815
Lower Regime								
T	36	—	219	—	225	—	220	245
SEE	0.00223	—	0.00197	—	0.00370	—	0.00145	0.00143
β_0	-0.004516	—	-0.000189	—	-0.001508	—	-0.000590	-0.000345
	(0.00383)	—	(0.00028)	—	(0.00095)	—	(0.00034)	(0.00029)
β_1	0.766	—	0.973	—	0.925	—	0.947	0.964
	(0.183)	—	(0.022)	—	(0.038)	—	(0.027)	(0.024)
Halflife	3	—	25	—	9	—	13	20
LST test	1.73*	—	1.41	—	9.37	—	1.70	3.92
(prob)	(0.188)	—	(0.495)	—	(0.009)	—	(0.427)	(0.141)
Tsay test	NA	—	2.03	—	10.09	—	0.99	3.45
(prob)	NA	—	(0.363)	—	(0.006)	—	(0.610)	(0.178)
Steady State	-0.0193	—	-0.0069	—	-0.0201	—	-0.0112	-0.0096
	(0.0020)	—	(0.0059)	—	(0.0039)	—	(0.0019)	(0.0028)
Convergence t-test	-0.76	—	-0.71	—	-1.34	—	-2.35	-1.31
Middle regime								
T	780	192	597	793	568	662	604	488
SEE	0.00149	0.00171	0.00106	0.00257	0.00195	0.00127	0.00099	0.00092
β_0	0.000036	0.000064	-0.000043	-0.000103	-0.000035	-0.000023	-0.000017	-0.000013
	(0.00006)	(0.00014)	(0.00006)	(0.00012)	(0.00010)	(0.00005)	(0.00004)	(0.00004)
β_1	0.992	0.993	0.996	0.981	0.994	0.982	0.990	1.012
	(0.005)	(0.016)	(0.007)	(0.008)	(0.013)	(0.008)	(0.008)	(0.011)
Halflife	84	98	183	37	111	39	68	NA
LST test	0.24	0.21	0.40	12.11	1.67	2.25	6.02	0.22
(prob)	(0.886)	(0.900)	(0.818)	(0.002)	(0.435)	(0.325)	(0.049)	(0.895)
Tsay test	0.33	0.67	0.38	13.05	2.68	2.79	5.70	0.94
(prob)	(0.846)	(0.714)	(0.826)	(0.001)	(0.262)	(0.248)	(0.058)	(0.624)
Steady State	0.0043	0.0091	-0.0114	-0.0054	-0.0055	-0.0013	-0.0017	0.0011
	(0.0065)	(0.0340)	(0.0342)	(0.0051)	(0.0136)	(0.0028)	(0.0048)	(0.0035)
Upper regime								
T	—	624	—	23	23	154	—	91
SEE	—	0.00108	—	0.00505	0.00505	0.00088	—	0.00122
β_0	—	0.000312	—	0.015875	0.015875	0.000692	—	0.000294
	—	(0.00016)	—	(0.00300)	(0.00300)	(0.00047)	—	(0.00074)
β_1	—	0.979	—	-1.077	-1.077	0.934	—	0.936
	—	(0.010)	—	(0.319)	(0.319)	(0.040)	—	(0.084)
Halflife	—	34	—	NA	NA	11	—	11
LST test	—	0.96	—	8.66	8.66	4.62	—	2.73
(prob)	—	(0.618)	—	(0.013)	(0.013)	(0.099)	—	(0.255)
Tsay test	—	1.21	—	16.89	16.89	2.26	—	3.48
(prob)	—	(0.546)	—	(0.000)	(0.000)	(0.324)	—	(0.176)
Steady State	—	0.0151	—	0.0076	0.0076	0.0104	—	0.0046
	—	(0.0021)	—	(0.0005)	(0.0005)	(0.0013)	—	(0.0058)
Convergence t-test	—	3.68	—	2.00	2.00	1.06	—	-0.39

Location and number of thresholds. It is clear that the nonlinearity problems are mitigated in the new model. The LST and Tsay tests generally confirm that there is no remaining nonlinearity in the subsamples after allowing for one threshold.[28] In only 3 subperiods (periods 3, 11, and 13) the LST and Tsay tests indicate the existence of remaining nonlinearity. For these cases we also present the estimated TAR(3;1,1) model. Inspection of the LST and Tsay tests show that this appears to solve the nonlinearity problem in the third and thirteenth period. In period 11 however there appears to be still significant nonlinearity in the lower and upper regime. For the remaining discussion we use the TAR(3) models for the 3 periods mentioned. (That is, we ignore the TAR(2) results in the table for these 3 periods.)

Convergence. Table 5 reports estimated steady states in each of the regimes. The theoretical model predicts that in the outer regimes convergence should be towards the thresholds. However, the estimation procedures from Table 5 did not impose this condition and thus we can use this as an additional test to evaluate the reasonableness of the theoretical model. Figure 5 plots the estimated thresholds versus the steady states and one observes a striking similarity between the two.

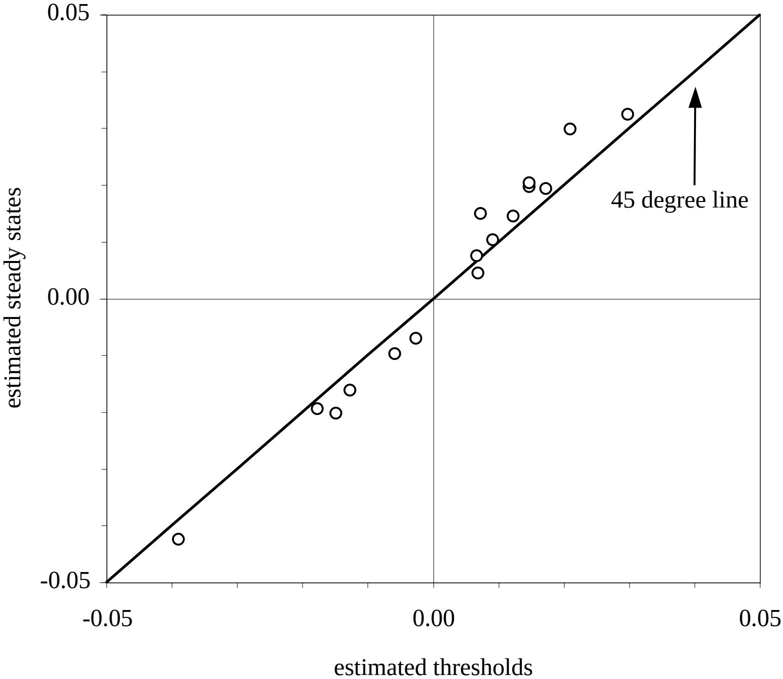
To construct a formal statistical test is again complicated because of the Davies problem. But if we take the thresholds as given it is trivial. In Table 5, we report test results as if the thresholds were known (in the row labeled "Convergence t-test"). Because the regression coefficients are well estimated the test for equality still often fails (7 out of 16 for a 5% level test, but only three for a 1% level test). However, we note that these tests do not take account of the uncertainty in the threshold estimates, so we feel encouraged by the results and consider this to be an independent confirmation of the model and the decline in the threshold over time.

The table also reports the estimated half-lives in each regime. We observe that convergence in the lower regime is fast—the half-life is always less than a month, and in four out of six cases less than 10 days. Convergence in the upper regime is slower than the lower regime and in only four cases it is less than a month. Convergence in the band is much slower than in the outer regimes, and $\hat{\beta}_1$ is usually within about 2 standard errors of unity (the sole exception is period 12). There are, of course, quite large standard errors on $\hat{\beta}_1$, and, since the coefficient is close to one, this would imply very large confidence intervals on the half-lives which are not reported here.

Declining thresholds. These results in Table 5 confirm our suspicion that the threshold declined remarkably over time. Figure 6 plots the absolute value of the threshold estimates γ_r together with a simple fitted linear trend over the thirteen periods. The absolute value of the thresholds decline from 0.0391 in the first period to 0.0059 in the last period, a reduction by more than a factor of six. The linear trend line shows a reduction by more than a factor of five. The oft-stated assumption that gold-points were

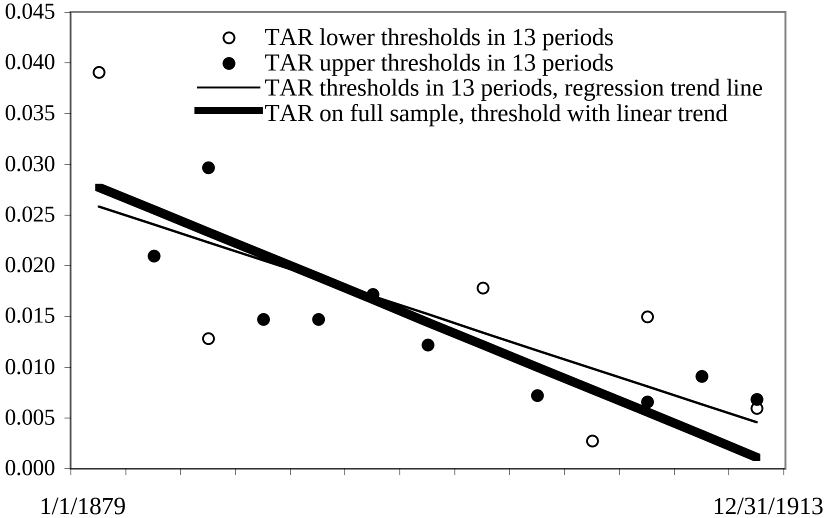
[28] Two adjustments have been made to these tests compared with the implementation in Table 3. First, the Augmented First Order LST test can not be computed in every regime because of multicollinearity due to limited number of observations. In these cases we computed the regular First Order test which does not have any third order terms. Second, the Tsay test can not always be computed if the data are put in ascending order, again because of multicollinearity. We present the maximum value of the Tsay test using both ascending and descending ordering of the data to minimize this problem.

Figure 5: Thresholds and Steady States: TAR Models in 13 Subperiods



Notes and Sources: See text and Table 5.

Figure 6: Evolution of Thresholds Over Time: TAR Models in 13 Subperiods



Notes and Sources: See text and Table 5.

Table 6: TAR(3) Model with Linear Developing Thresholds

T		10616	
log L		51452.7	
SC		-9.6786	
Threshold par α_0		0.028859863	
α_1		-2.7185E-06	
Regime	Lower	Middle	Upper
T	1766	5095	3755
SEE	0.0024	0.0022	0.0014
β_1	1.147	1.062	1.082
	(0.0257)	(0.0138)	(0.0158)
β_2	-0.153	-0.015	-0.029
	(0.039)	(0.021)	(0.022)
β_3	-0.022	-0.024	-0.060
	(0.038)	(0.021)	(0.021)
β_4	-0.002	-0.030	-0.005
	(0.026)	(0.014)	(0.015)
halflife	25	98	61

Notes: See text and footnote 29 for details of the estimation procedure. Standard errors in parentheses.

pretty much stable in the entire 1879–1913 period would appear to be refuted by our analysis. We also note that in the results in Table 5 and Figure 6 there does not appear to be any striking asymmetry in the thresholds. We have no reason to doubt that the gold points were nearly symmetric during the sample period.

Figure 6 also displays an alternative way of modeling evolving thresholds, that is, by making the γ_r depend linearly on time in an explicit fashion. Of course, it is then necessary to also model the point of convergence as dependent on time. To make the estimation feasible, we impose the restriction that the thresholds are symmetric and that outer-regime convergence is always to the threshold as predicted by the theory and as seen above. This reduces the search to a two-dimensional parameter space.[29]

We found the best model to have 4 lags and a delay parameter of 4. The results are reported in Table 6, and the estimated absolute threshold size is plotted in Figure 6 as the heavier solid line. This line is remarkably close to the linear regression line estimated from the thresholds in the 13 subperiods, and also has a marked downward trend.

[29] Thus, we estimate the TAR($R; k, d$) of equation 8 with $R = 3$ and with the restrictions

$$\gamma_1(t) = -\gamma_2(t), \quad \beta_0^1(t)\left(1 - \sum_{i=1}^{k} \beta_i^1\right)^{-1} = \gamma_1(t), \quad \beta_0^3(t)\left(1 - \sum_{i=1}^{k} \beta_i^3\right)^{-1} = \gamma_2(t),$$

where $\gamma_2(t) = \alpha_0 + \alpha_1 t$. The values of $\hat{\alpha}_0$ and $\hat{\alpha}_1$ are obtained by a simple grid search over an increasingly refined grid. The number of lags k and the delay parameter d were chosen to minimize SC from the same set as in Table 3.

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, using our new sources, we can now present a limited quantity-based cross-check on our method in Figure 7 for the time period for which we have Mint report data on gold exports. There we plot, at daily frequency : the exchange rate and parity, our gold export point estimates, and Officer's GPA export gold point. (GPA denotes gold point arbitrage.) The circle markers indicate, on the exchange rate plot, those days on which substantial gold exports (more than $50,000) were observed by the Mint. To account for the short—but possibly significant—lag in the buying of bills and the shipment of gold, we plot the maximum of the exchange rate in the three days before the actual shipment. At the bottom of the graph we also show the actual volume of the exports reported by the Mint.

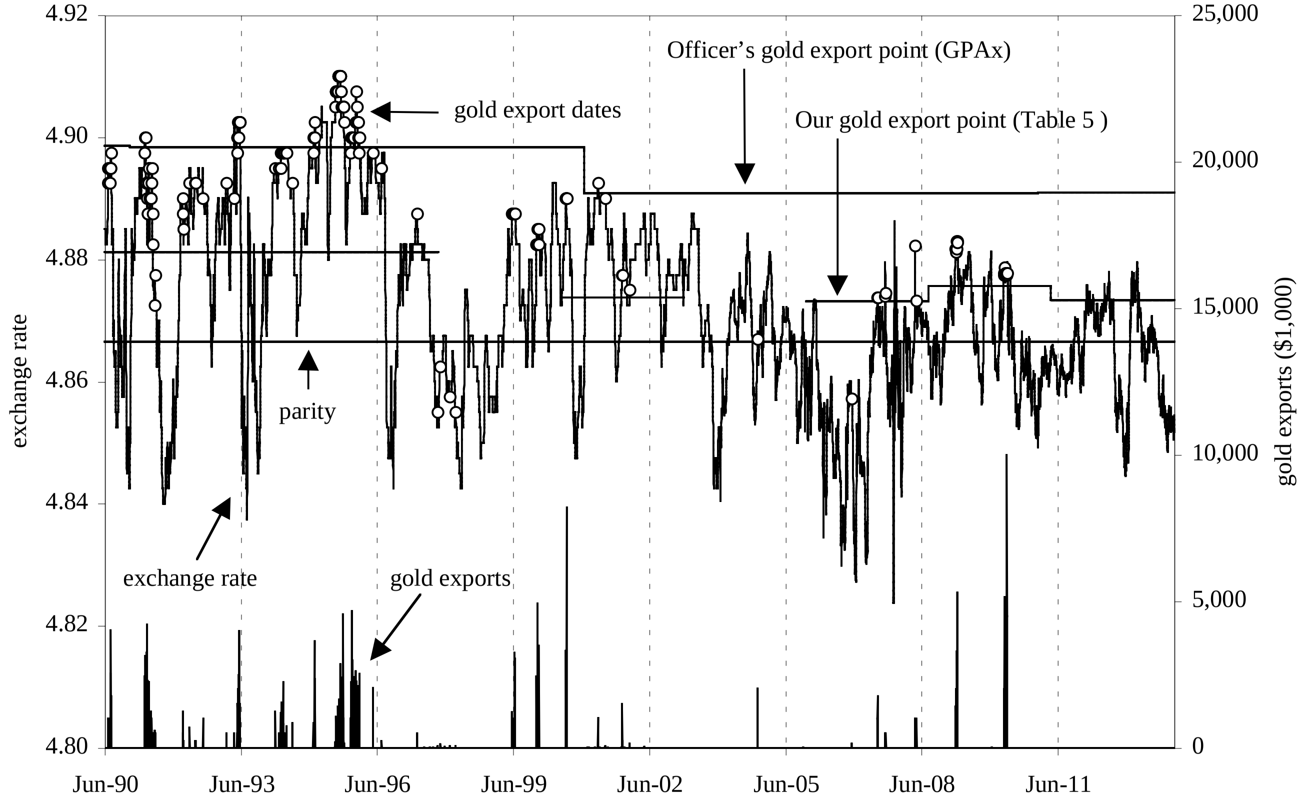
Do the data look reasonable? As mentioned before gold flow data are not very reliable. This is a possible reason for the five observations where gold was shipped to the U.K. when the exchange rate was actually below parity. It might also be the case that for these observations gold was shipped for other reasons than gold point arbitrage. This explanation seems to be favored by the fact that the volume shipped was 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.

The figure sheds light on where the true gold points were. Our gold point estimates predict actual gold flow quite well for the period 1890–96. For this period we observe almost no gold movement when the exchange rate was within our estimated band, and we observe gold flow when the exchange rate is above our estimated gold point almost every time. In fact, during this period 98 percent of the gold shipped ($224 out of $227 million) was shipped at times when the exchange rate was above our estimated gold point, neglecting periods when such an estimate was unavailable. Arguably, the gold point might be slightly higher than our estimate, but Officer's GPA gold export point is almost certainly too high, except possibly for a short period in 1895, since we find that only 34 percent of gold flows ($87 million) actually occurred when the exchange rate was above his estimate.

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. At the very end, in 1912 and 1913 the graph suggests that our gold point estimate is a too low as no gold flows are observed while one would have predicted some based on our gold point estimates.

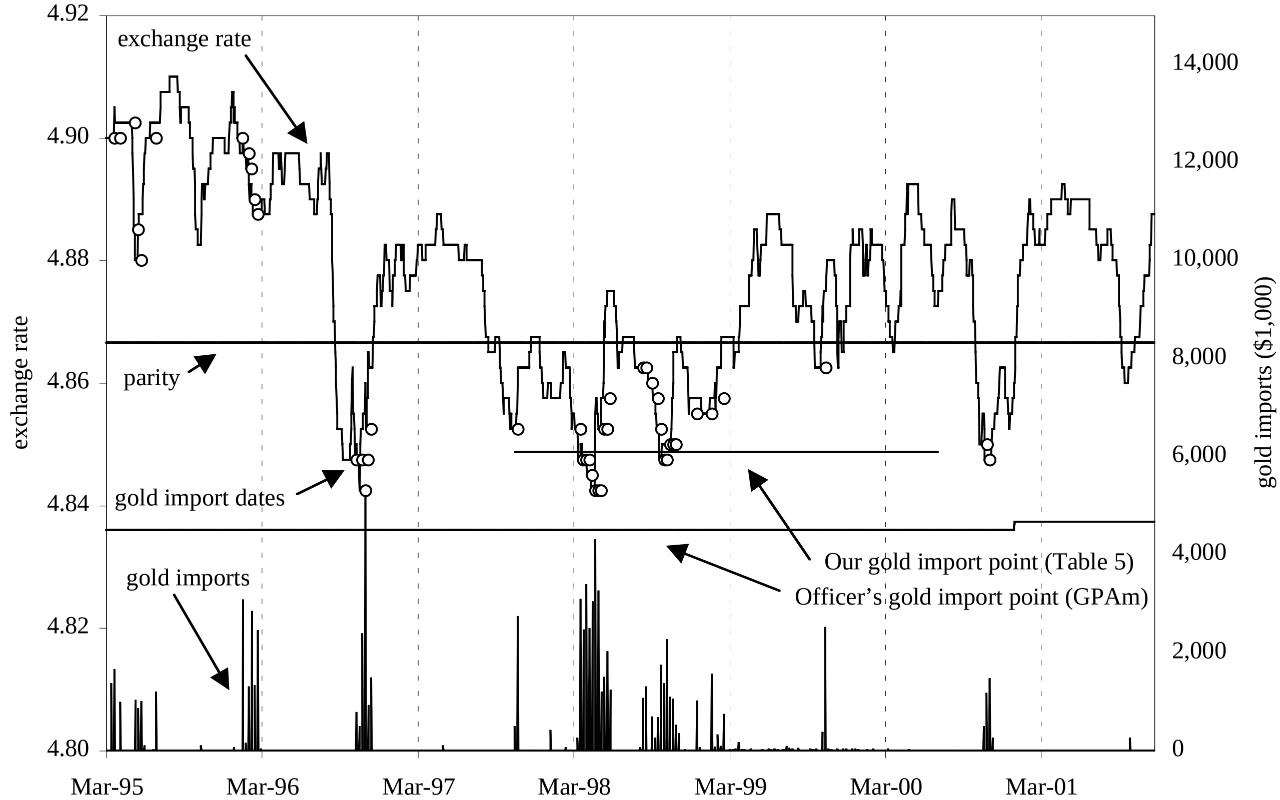
Using the same approach we can examine arbitrage in the opposite direction, although the data here are more problematic. Figure 8 shows the same kind of graph but for gold imports, employing Coleman's gold flow data for 1895 to 1902 and the estimated gold import points. About $ 86 million in gold imports took place over the entire

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.

period. In the figure, only imports of more than $350,000 in a week are considered to be substantial, and they are indicated with circles.

There are several reasons why this graph is less reliable than Figure 7. First, Coleman only collected weekly aggregates of gold imports; hence the timing of the gold imports is less precise and for this reason we have plotted the minimum exchange rate over a span of 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. Some of these quality issues are apparent in the import data in 1895 and 1896, where large imports are observed at an exchange rate far above parity.

This begs the question as to what we should make of the large gold imports observed in the beginning of 1899 at exchange rates just below 4.86. If these data are correct then our estimated gold points seem too high (that is, too far from parity). However, if these observations are invalid then it appears that our estimates seem too be hitting the revealed gold export points almost exactly. Judged on the same basis, it would seem that Officer's GPA gold import point is almost certainly too high, since it would have predicted no gold flows at all in this period.

5 Conclusions

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 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 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."[30] 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 versus 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 margins had certainly been shrinking. Perhaps as a result of competition increasing among banks. The correspondent for *Bankers Magazine*, could report on June 1910 (p. 924) that "there is not much profit in shipping gold as might be thought, a thousand dollars on each million being considered quite enough of 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.

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

References

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

Balke, N. S., and T. B. Fomby. "Threshold Cointegration." *International Economic Review* 38 (August 1997): 627–46.

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."' Princeton, N.J., March 1998. Photocopy.

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

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

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

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: 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.

Gallant, A. R., P. E. Rossi, and G. Tauchen. "Nonlinear Dynamic Structures." *Econometrica* 61 (July 1993): 871–907.

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. "Inference When a Nuisance Parameter is Not Identified Under the Null Hypothesis." *Econometrica* 64 (March 1996): 413–30.

Hansen, B. E. "Inference in TAR Models." *Studies in Nonlinear Dynamics and Econometrics* 2 (April 1997):1–14.

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

Hansen, B. E. "Sample Splitting and Threshold Estimation." *Econometrica* 68 (May 2000): 575–603.

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

Luukkonen, R., P. Saikkonen, T. Teräsvirta. "Testing Linearity Against Smooth Transition Autoregressive Models." *Biometrika* 75 (September 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.

O’Connell, P. G. J., and S.-J. Wei. “ ‘The Bigger They Are, The Harder They Fall’: How Price Differences Across U.S. Cities Are Arbitraged.” Working Paper Series no. 6089, National Bureau of Economic Research, July 1997.

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, 1998.

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 1925–1931 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.

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

Potter, S. M. “Nonlinear Impulse Response Functions.” Staff Report no. 65, Federal Reserve Bank of New York, February 1999b. Forthcoming in *Journal of Economic Dynamics and Control*

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.

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 (August 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 (March 1989): 231–240.