 Oxford Economic Papers 52 (2000), 745–773 **745**

Taxing and spending in the long view: the causal structure of US fiscal policy, 1791–1913

By Kevin D. Hoover* and Mark V. Siegler†

* Department of Economics, University of California, Davis, California 95616–8578, USA; e-mail: kdhoover@ucdavis.edu
† Department of Economics, Williams College, Williamstown, MA 01267, USA; e-mail: Mark.V.Siegler@williams.edu

Causal relations between US federal taxation and expenditure are analyzed using an approach based on the invariance of econometric relationships in the face of structural interventions. Institutional evidence for interventions or changes of regime and econometric tests for structural breaks are used to investigate the relative stability of conditional and marginal probability distributions for each variable. The patterns of stability are the products of the underlying causal order. Consistent with earlier work on the post World War II period, we find that dominant causal direction (with only a short-lived reversal) runs from taxes to spending in the period before World War I.

1. Tax and spend or spend and tax

In many countries, the state of fiscal policy and, in particular, the size of the government's budget deficit is an important political, as well as economic, issue. In thinking about how to bring deficits under control, implicit assumptions are made about the normal workings of the political and administrative processes that determine fiscal policy. There are at least two views of the relationship between the revenue and expenditure sides of the government's budget. In one view, spending causes taxes in the sense that, if spending is increased, the normal response (without any presumption of institutional change) is for taxes to be brought into line with spending. Such a mechanism has been stigmatized in the United States with the phrase that the politicians who insure such realignment of taxation with expenditure act as 'tax collectors for the welfare state'. On the other view, taxes cause spending in the sense that if tax revenues are found to be higher, the normal response is for new spending to absorb the new revenues (rather than tax rates to be cut or debt to be repaid). On the first view, those who advocate political reforms aimed at shrinking the size of the government should aim directly at the control of expenditures. On the second view, the limitation of tax revenues would be an adequate method of achieving their goals. The issue is one of the causal structure of the budgetary process. In this paper, we examine the historical causal relation between taxes and spending in the United States, using data from 1791 to 1913.[1]

We define causal order according to the definitions of Herbert Simon (1953). Simon's analysis of causal order is well-known to econometricians, but empirical inference grounded in his analysis is less familiar. Hoover (1990) developed a general methodology for empirical causal assessment consistent with Simon's analysis, and Hoover (1991) applied the methodology to the relationship between money and prices. The current paper more directly extends and develops the analysis of Hoover and Sheffrin (1992), which considered the fiscal history of the United States after the Second World War. It represents a clear improvement over that article in several respects.

One goal of this paper is to apply the general methodology in a context in which some of the limitations that made the data used in the earlier study difficult to interpret no longer bind so tightly. First, the earlier study used a relatively short sample. This might not seem obvious, because the data were quarterly. The Federal budget cycle is, however, annual, and sampling quarterly does not seem calculated to add four times the information as it would in some cases. The current study uses annual data over a much longer period.

Second, in order to isolate discretionary spending and to avoid the relationships between taxes and spending that result from the so-called 'automatic stabilizers' (i.e., countercyclical transfer payments and 'tax-bracket creep'), the data in Hoover and Sheffrin (1992) were pre-filtered to remove the cycle and the effects of inflation. There is some risk in this strategy that the relationships discovered between taxes and spending may, to some extent, be artifacts of the filtering procedure. The current study confronts these two problems directly. It focuses on over a hundred years of fiscal behavior, and uses data that have not been pre-filtered.

In principle there might be losses from not accounting for automatic stabilizers, but in practice it is unlikely to matter. The United States did not have a permanent income tax until 1913. In addition, Federal unemployment insurance and welfare payments remained extremely small before the passage of the Social Security Act in 1935. Lindert (1994, Table 1B, p. 11) estimates that government spending on welfare and unemployment relief as a percentage of national product was less than 0.5% through 1930. Any neglected effect of automatic stabilizers is too small to matter for the results reported below.

A long run of data is useful in its own right, but more so because another goal of the current study is to extend the span of the history considered, to determine whether the qualitative conclusions of the earlier study are robust to a wider range of institutional variation. On a first pass, we consider 200 years of fiscal policy. It turns out, however, that for reasons that are discussed in Section 5 below, data from the two world wars and the interwar period cannot be given a straightforward

[1] As discussed below, we initially examined data from 1791 to 1990, but the institutional history of the middle 20th century turns out to limit the informativeness of the data in that period.

causal interpretation. As a result, although we have in fact applied the methods outlined in this paper to 200 years of US fiscal data, to save space and the reader's patience, we report only from the period before the First World War.

Yet another goal is to set some bounds—at least for particular historical periods—on the sorts of detailed models of the fiscal process that are admissible. The causal knowledge gained from such a study is coarse. Nonetheless, because particular models imply definite causal orderings, to the extent that the data support one causal order, models incompatible with that order are not supported. A study such as this will not point to a particular detailed model, but may provide evidence for or against a class of models that incorporates a particular causal order.

Finally, the Hoover and Sheffrin (1992) paper reached some substantial conclusions which may or may not be robust. In that article, it was argued that the best interpretation of the post-war data was that for two decades taxes caused spending. The most surprising finding was that sometime in the late 1960s or the early 1970s, the causal order changed completely. After 1970, taxes and spending were causally independent. In retrospect, the change in causal order in itself should not have been surprising. Unlike the causal relations between inanimate objects described in physics, economic relationships are the products of human institutions; and institutions change. In some cases, we might expect that the causal relationships between aggregate data are the product of the complex, but overall fairly stable, interactions of many economic actors. The relationship between money and prices (cf. Hoover, 1991) or between consumption and income might well be of that nature. In such cases, causal direction might be one way or the other, but one would not necessarily expect it to change qualitatively over time. In the case of fiscal policy, in contrast, taxes and spending are set by the complex interaction of a much smaller number of actors: the President and his staff, the Congress and its staff. Thus, while it is believable that no single actor controls the budget in the sense of being able to directly and independently choose the dollar amounts, it is more reasonable to suppose that, from time to time, wholesale reform, change, or realignment of these processes is possible in a way that would appear to change the causal order between them. And, indeed, in the case in question, political scientists had argued that, in fact, there had been a sea-change in the conduct of fiscal affairs in the early 1970s—precisely at the time suggested by the data (see Wildavsky, 1988, p. 120).[2] That causal order is malleable is independently interesting, yet it in no way detracts from the utility of knowing the true causal order at a given time. For it is that order that must be used to achieve budgetary goals.

[2] The causal methodology employed here relies on a coordination of statistical evidence with information from narrative histories of the fiscal policy. Here these histories are used instrumentally to provide corroborating evidence for the periods of stability and the occurrences of well-identified interventions in the processes that determine taxes or spending. The history is in fact interesting in its own right and a fuller historical analysis based on this causal study is presented in the earlier version of this paper (available from the authors on request).

2. Causality in the budget process

Taxes and spending relative to potential output are displayed in Fig. 1 on a logarithmic scale. The coarser outlines of US fiscal history are immediately evident. Wars have typically been financed through deficits; but wars are episodic. In the interbellum periods, the bulk of the 200 years pictured in Fig. 1, revenues exceeded expenditures and the national debt was paid down. This pattern is broken only in the past quarter century. Aside from wars, Federal expenditures were low until the 1920s. The New Deal and the Second World War carried the Federal expenditures to new highs, peaking at about 40% of potential GNP. Expenditures fell back substantially from these peaks, and have remained untypically stable at about 15% of potential GNP since the end of the Korean War. What is more, expenditures and taxes are much better aligned in the post-Korean War period than in any similar period in history: neither surpluses nor deficits are very large percentages of potential GNP.[3]

How can these data be interpreted causally? There are two issues. First, what does it mean for one variable to cause another in the control sense? Second, how could one infer the true causal order? In Section 3 below, we briefly sketch the main lines of an empirical methodology that addresses these issues (fuller expositions are found in Hoover, 1991, and Hoover and Sheffrin, 1992).

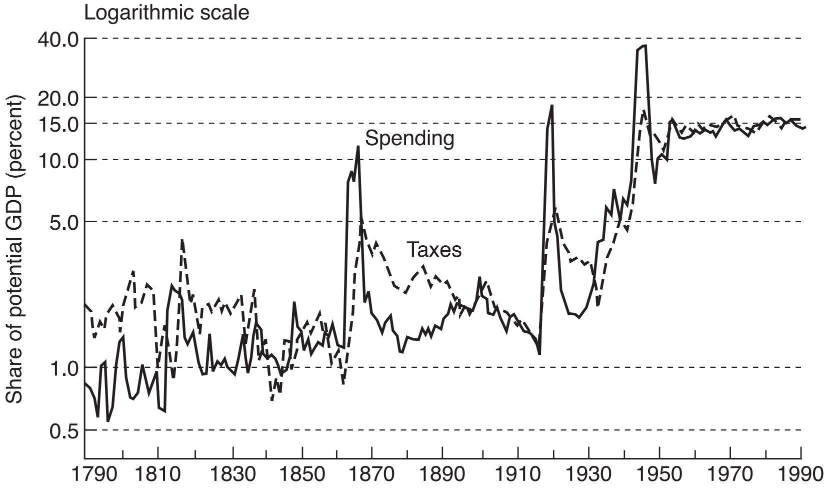

Fig. 1. US federal taxes and spending, 1790–1990

[3] The observant reader will note that taxes exceed spending for many of the later years in Fig. 1 despite the existence of large deficits. The reason is that Fig. 1 plots expenditures net of interest payments, because we are principally interested in discretionary expenditures, and interest payments have often been greater than the gross deficit.

Simon (1953) provides a satisfactory analysis of the first issue.[4] For Simon, X causes Y means that in the true process that generated the data, X is recursively ordered ahead of Y, so that one can indirectly control Y through any means that allow one (directly or indirectly) to control X.[5] In Barro's (1979) tax-smoothing model, spending causes taxes. Given exogenously formed spending plans, the government sets tax rates in order to equalize the marginal cost of tax collection over time. Spending is recursively ordered ahead of taxes: an intervention that alters spending plans (an independent random shock or a change in the parameters governing the time-series properties of spending) is transmitted in the Barro model to changes in tax rates; while an intervention that alters tax rates through some other mechanism than spending itself (a different independent random shock or a change in a parameter that does not also directly govern spending) is not transmitted back to spending.[6]

Because there are alternative plausible theoretical models with different causal orderings implied in their structures, causal direction cannot be determined *a priori* simply by writing down a theory. Barro motivates his model with the observation that taxes are distortionary. Government spending itself might be distortionary; and, if it is, contrary to Barro's assumption, an optimizing government might take its revenue path as a constraint and arrange its spending plans to equalize the marginal cost of these distortions: taxes cause spending. Once it is recognized that there can be costs on both sides of the balance sheet, then a model in which tax and spending plans are made simultaneously is possible: causality is mutual—taxes cause spending, and spending causes taxes. It is also possible that taxes and spending are not set with any reference to the other. In this case, the two would be causally independent, and the deficit would be an accident of history.[7]

[4] See also Simon and Iwasaki (1988). Hoover (1990) relates Simon's analysis to a wider philosophical literature on causality.

[5] In the linear case analyzed by Simon, a model is causally ordered if and only if it is identified. In empirical investigations, we do not know the true structure, so the correspondence of identification and causality cannot be exploited inferentially.

[6] In an earlier version of this paper, Hoover and Siegler (1997), (available from the authors on request) provide a detailed analysis of both the causal ordering of the Barro model and the problem of causal inference when it is *ex hypothesi* the true process that generates observable data.

[7] Causal independence is possible only if the government's intertemporal budget constraint is not operative. Although the period-by-period government budget constraint (any deficits must be financed by debt or monetized) certainly holds, the intertemporal budget is a hypothesis about the forward-looking behavior of the government. It is often assumed in theoretical models in order to cast them as proper optimization problems. Nonetheless, it is an empirical question whether the government behaves as if it must balance its budget over an infinite horizon, and equally an empirical question whether other agents in the economy incorporate this assumption about the government into their own decisions (see Trehan and Walsh, 1988). What is more, as the analysis of the Barro model in Hoover and Siegler (1997) demonstrates, the methods employed in this paper are informative whether or not the government's intertemporal budget constraint is operative.

An important lesson of the interpretation of causality as controllability is that there are no easy answers to the second question, how could one infer causal order? There is a natural temptation to look at Fig. 1 and conclude that the question of causal direction can be settled by visual inspection. Taxes appear to lag behind spending, so one is tempted to say that spending causes taxes. One must resist this temptation; it would be an example of the ancient fallacy *post hoc ergo propter hoc*. Similarly, one sees periods of war and makes the correct observation that tax interventions did not cause the war and, therefore, the war spending; and so spending must cause taxes. That too would be an unwarranted. Facile conclusions based on simple descriptions of the data as in Fig. 1 are confused on our understanding of causality, because causality is a question of the structural relationship between variables and not a question of the particular path of realizations of those variables.

Revenue changes do not generally cause wars—true enough. But spending might have multiple causes. The relevant question is, given that a war causes an increase in spending, does the increase in revenues that follows arise as a matter of course, because the true process has a particular recursive structure and a particular set of deep parameters, or does the increase in revenues follow only because those deep parameters have been changed by deliberate changes in policy? In the first case, one would be justified in saying that spending causes taxes. In the second case, the recursive structure of the fiscal process may still order taxes ahead of spending (i.e., taxes cause spending). In that case, however, the change in parameters shows up as a distinct pattern of instabilities in the statistical characterizations of the relationships between revenues and spending. As we show in Section 3 below, such instabilities can be used to reveal the causal structure of the underlying fiscal processes, which cannot be discerned in stable policy regimes because of observational equivalence.

The point of our investigation is not to characterize the extraordinary events of US fiscal history, such as wars, but to use those extraordinary events to learn what the normal causal structure of the fiscal process is in fact. The difficulty with Fig. 1, despite the temptation to provide a direct causal interpretation, is that there will always be some set of conceivable interventions that makes any particular causal model deliver the observed path (the problem of observational equivalence). The inferential problem, which requires deeper investigation, is to determine what causal structure (and therefore what class of particular models) is compatible with the actual interventions present in the historical record.

One must also resist the temptation to interpret the data through econometric procedures that are the functional equivalent of Fig. 1. The sense of causality here is different from Granger (or statistical) causality familiar to most economists: Y Granger-causes X if Y and its past help to predict X conditional on its own past. Granger-causality is matter of forecastability or predictability, but Granger (1980, p.351) himself observes ‘that controlability [*sic*] is a much deeper property than [Granger] causality, in my opinion, although some writers have confused the two concepts. If Y [Granger] causes X, it does not necessarily mean that Y can be used to control X'.[8]

3. The identification of causal order

To better understand the problem of causal inference in economic data, consider a simplified example. Define a description of the world as consisting of three elements: parameters, which may be selected mutually independently by economic agents or nature; random shocks, which are white noise innovations; and variables, which are the product of the interaction of parameters and random shocks, and which are at best indirectly controllable through control over parameters.

Now consider two data-generating processes:

Process I

$$R = a + \varepsilon \tag{1}$$

$$X = bR + \nu \tag{2}$$

Process II

$$X = d + \omega \tag{3}$$

$$R = eX + \zeta \tag{4}$$

where X is spending, R is taxes, lower case Roman letters are parameters, and lower case Greek letters are random shocks (white-noise innovations). Using Simon's (1953) recursiveness criterion, taxes cause spending in Process I, while spending causes taxes in Process II. Without knowledge of the parameters one cannot distinguish whether either of these causal orders or some other order is correct: The two processes have precisely the same reduced forms. This is the famous problem of observational equivalence (cf. Basmann 1965, 1988; Neftçti and Sargent 1978).[9]

In the face of observational equivalence, how might we establish which causal order is in fact correct? To take a simple example, if the innovations are all independently, normally distributed, then eqs (1)–(4) can be interpreted as regression equations. Suppose that Process I is in fact the true causal order, and that we know from historical narrative or other independent information that there is a regime change in the process that governs taxes. This would be represented by a change in the parameter a or in the variance of σ_{ε}^2. Such an intervention would therefore alter the estimated coefficients of a regression of R on a constant; but it would not affect the coefficients of X on a constant and R, which would be governed by the parameters of eq. (2). This is because these regressions, in effect, recapitulate the causal structure of the world. But what if we reversed

[8] Jacobs *et al.* (1979) and Engle *et al.* (1983) provide detailed analyses of the relationship of Granger-causality to control notions of causality; also see Hoover (1988, ch.8).

[9] The example here is simple, involving only contemporaneously related variables, but the problem generalizes, for example, to vector autoregressions (cf. Cooley and LeRoy, 1985).

these regressions. Suppose that we wrongly thought the causal order was as in Process II. Consider a regression of X on a constant

$$X = f + \xi \tag{5}$$

Substituting eq. (1) into eq. (2) we obtain

$$X = ab + b\varepsilon + \nu \tag{6}$$

The constant in eq. (5) is $f = ab$; and the variance of the error term is $\sigma_\xi^2 = \mathrm{var}\,(b\varepsilon + \nu) = b^2\sigma_\varepsilon^2 + \sigma_\nu^2$. The intervention in the tax process would evidently cause the estimated coefficients f or σ_ξ^2 to change; that is, eq. (5), unlike eq. (2), is not invariant to changes in the structure of the tax process. The algebra is somewhat more complicated, but it is also easy to show (cf. Hoover 1990, p.225; 1991, p.385; or Hoover and Sheffrin 1992, p.230) that when Process I is the true process, the coefficients of

$$R = g + hX + \eta \tag{7}$$

will not be invariant to changes in the process governing spending.

The importance of instability and invariance in establishing causal order has been explained with reference to simple processes in which the underlying structure is known. In general, we do not observe the underlying structure directly. We can, however, estimate probability distributions for the observable variables. If we denote the joint distribution of the variables, in this bivariate case, D(R,X), then the problem of observational equivalence corresponds to the fact that a joint distribution can be partitioned in two ways into the product of a conditional and a marginal distribution: we can call relationships like eqs (2), (4), and (7), in which both variables appear, conditional regressions (or distributions); while those like eqs (1), (3), and (5) can be called marginal regressions (or distributions). Thus, eq. (2) defines the distribution of X conditional on R; while eq. (1) defines the distribution of R marginal of X. These relationships are not structural, but correspond to reduced-form regressions. Although we have concentrated on the bivariate case for expositional purposes, the analysis naturally generalizes to multivariate cases. These are important to the empirical implementation later in the paper in which we consider dynamic forms with lagged dependent and independent variables. If we are able to observe econometrically the invariance or lack or invariance of the different conditional and marginal regressions in the face of interventions that are associated with either the tax or the spending processes, then the pattern of invariance will reveal the causal order that must have generated the data.

Table 1 summarizes the invariance patterns of the simple bivariate example above in the form of an decision algorithm. We seek to determine the causal order between X and R. Let the suffix C indicate a conditional regression and the suffix M a marginal regression. There are two steps:

Stage (I) Identify the nature of the break. If both the conditional and the marginal regressions for X (XC and XM) indicate a break statistically and if the narrative evidence supports an intervention in the X process, we identify the break as an intervention in X and proceed to the second stage. Similarly, if both RC and RM

Table 1 An algorithm for causal inference

Stage 1: Identification

A. If (i) There is narrative evidence of an intervention in the X process; and
(ii) Both XC and XM break statistically,

Then, there is a genuine intervention in the X process: GO TO STAGE 2.A BELOW.

B. If (i) There is narrative evidence of an intervention in the R process; and
(ii) Both RC and RM break statistically,

Then, there is a genuine intervention in the R process: GO TO STAGE 2.B BELOW.

Stage 2: Discrimination

A. Interventions in the X Process

RM

RC	Stable	Unstable
Stable	$R \perp X$	$X \Rightarrow R$
Unstable	$R \Rightarrow X$	$X \Leftrightarrow R$

B. Interventions in the R Process

XM

XC	Stable	Unstable
Stable	$X \perp R$	$R \Rightarrow X$
Unstable	$X \Rightarrow R$	$R \Leftrightarrow X$

Key: $\Rightarrow$ = "causes," $\Leftrightarrow$ = "mutually causes," $\perp$ = "is causally independent of."

break and the narrative evidence supports an intervention in the *R* process, we identify the break as an intervention in *R* and proceed to the second stage.

Stage (II) Discriminate between possible orders. If we identified an *X* break, we can use the two-by-two table in II.A. It shows the possible patterns of breaks in the *R* process and identifies each with a causal order conditional on having previously ascertained that the break was in *X* process. So, for example, a pattern of RC stable and RM unstable is consistent with *X* causing *R*. The table shows all the possibilities. If we identified an *R* break in the first stage, we can use *mutatis mutandis* the similar two-by-two table in II.B.[10]

The existence of cross-equation restrictions in the data-generating process of the type that are often associated with rational expectations complicates the interpretation of the instability of conditional distributions (see Hoover, 1990, appendix; Hoover, 1991, pp. 386–7; and Hoover and Sheffrin, 1992, pp. 230–1). The essential difficulty is that, if *A* causes *B* and there are cross-equation restrictions, then an intervention in the process for *A* will not, as suggested in Table 1, leave the conditional distribution of *B* invariant. It is nevertheless true that an intervention in the process for *B* will leave the marginal distribution for *A* invariant. It is, therefore, essential in such cases to have a rich menu of interventions in both processes.[11] In practice, however, these problems do not appear to be relevant to the current paper, as we found no evidence in our data of these sort of complications. We eschew further discussion of them to simplify the presentation of the methodology for causal inference.

The general strategy of this paper is then straightforward. We use the historical narrative to identify periods in which no interventions in either taxes or spending occur—i.e., where their relationship is able to work in its usual manner—as well as times when interventions occur on one side of the relationship or the other. We use the tranquil periods as baselines over which to specify particular representations of the two marginal and two conditional regressions. We then test whether each of these regressions breaks down at a point identified from the narrative as a possible intervention in one of the processes. The pattern of breaks is then analyzed according to the algorithm in Table 1.

[10] Table 1 is incomplete. If both conditions A and B in Stage I are fulfilled then causal discrimination is not possible. In such a case, in which every regression breaks and the narrative supports interventions on both sides, it might be that causation is one-way in either direction, that causation is mutual or that there is causal independence there is simply no way to tell. This problem becomes practically important in the 20th century (see Section 5 below).

[11] Hoover and Siegler (1997, Appendix 2) provide a detailed example of the problem of cross-equation restrictions, using the tax smoothing model as an illustration. The emphasis on stability calls to mind Engle *et al.*'s (1983) concept of strong exogeneity. Evidence of strong exogeneity can be used in the right circumstances to help infer causal direction. The existence of cross-equations restrictions, however, presents cases in which causal order is well defined even though the cause is not strongly exogenous for the effect. The relationship of causality to exogeneity is discussed at length in Hoover (2001).

4. Stability and change

4.1 The data

The sources and definitions of the data used in this study are described in detail in the Appendix. The two basic series are the current and lagged values of real Federal government expenditures net of interest payments as a share of potential real GNP (X) and real Federal government receipts as a share of real potential GNP (R). We focus on expenditure net of interest because we are interested in the discretionary actions of the fiscal authorities. The series are scaled by potential GNP to account for the growth of the economy over 200 years and to render them stationary.[12] Even though the data are not filtered as in Hoover and Sheffrin (1992) to adjust for cyclical fluctuations, scaling by potential GNP reduces adventitious fluctuations in tax and expenditure shares due to unexpected or cyclical fluctuations in output.

The Federal fiscal data are good quality. Questions might be raised about the quality of the GNP and price data. These concern us less in the current context than they might if, for example, we were largely concerned with the cyclical behavior of output. First, for the period considered in this paper, they are the only usable data we know of. Second, they are used only for construction of the potential GNP series. The point of scaling by potential output is to provide a measure of what share of the economy is taken by the Federal government abstracting from cyclical variation. Any series that approximates the low frequency shape of real GNP would probably serve that function equally well. The quality of the price index is not likely to be critical as, to a first approximation, it cancels out in the calculation of X and R. The principal inaccuracy of the GNP data arises from its interpolation between benchmark years. But this is unlikely to matter in this context as our potential GNP series smoothes variations between business-cycle peaks and, therefore, does not use the questionable interpolated elements of the GNP series to any great extent.

4.2 Chronology of the tax and spending processes

It is the nature of government that changes in taxing and spending plans are made constantly. Not every such change should be regarded as an intervention in the underlying causal processes that connect taxes to spending. One of the central difficulties in executing the strategy of causal inference outlined here is that some judgement must be made as to which events are significant, and on which process they operate. This is one area in which the statistical tools provide a useful cross-check. It is necessary in advance to identify some periods as tranquil and some periods as containing interventions on one side or the other, but econometric

[12] A glance at Fig. 1 should convince the reader that either the data are non-stationary, even after scaling by potential output, or (and we find this more likely) they come from distinct regimes with distinct means. Any of our tranquil periods (see Section 4.2 below) shows data that appear not to have a common mean without trend or drift. Formal tests of stationarity are known to suffer severe lack of power in short samples, so we have not attempted to compute them for the tranquil periods. In any case, our dynamic forms are general enough to allow for cointegrated data.

tests of structural stability can be used to confirm the identifications. The relationship is mutual. Regressions might show apparent structural breaks because they are misspecified. Statistically identified breaks will be regarded as genuine and not artifacts of misspecification only if they correspond to known historical events of the right type.

Table 2 presents a selected chronology of the taxing and spending processes for two centuries. It identifies events that serve as historical benchmarks, and events that are potentially structural interventions of the type needed to identify causal order. It is a rich list, although it leaves out a large number of events that we believed were of secondary importance.

The first use of Table 2 is to identify tranquil periods on which to specify baseline marginal and conditional regressions.[13] We examine six possibly tranquil periods: (1) 1791–1806; (2) 1816–28; (3) 1848–60; (4) 1872–84; (5) 1900–12; (6) 1954–63.[14]

4.3 Baseline regressions

For each tranquil period, four regressions were estimated corresponding to the conditional and marginal regressions of Section 3 above for revenues (R) and expenditures (X). To obtain a well specified regression, we apply the general-to-specific modelling framework of David Hendry.[15]

We begin with an unrestricted distributed lag regression of each variable. For the conditional regressions, the regressors are a constant, two lags of the dependent variable and the current value and two lagged values of the other variable. For the marginal regression, the current and lagged value of the other variable are omitted. With a few exceptions we were able to find general forms for which we could not reject the nulls of normality, no serial correlation up to second order, and no first-order autoregressive conditional heteroscedasticity. There are, however, a few exceptions. Typically, one would want to add further lags to the regressors or expand the set of regressors in order to eliminate these indications of misspecification. Unfortunately, the tranquil periods are often short, which limits the degrees of freedom. Given that constraint, the unrestricted forms are well specified.[16]

[13] Tranquil periods were determined from a careful reading of the historical record. As a cross check, their statistical tranquillity is checked through tests of the within-sample stability of the parsimonious forms and through the recursive regressions also used in out-of-sample stability testing. See Sections 4.3 and 4.4 below.

[14] 1974–79 is also a potential tranquil period and was so identified in Hoover and Sheffrin (1992), but it is too short to support the specification of the necessary regressions, so we do not treat it as one here. As mentioned in Section 2 above (discussed further below), the world wars and the interwar period show too many structural breaks that simultaneously affect both tax and spending processes to be useful for causal discrimination. Results for period (6) are therefore not reported in this version of the paper.

[15] See Hendry and Richard (1982), Hendry (1983, 1987). Also see McAleer *et al.* (1985), Gilbert (1986), Pagan (1987), Phillips (1988), and Mizon (1995) for general defenses of this methodology.

[16] The summary statistics from these regressions are reported in Table 3 of Hoover and Siegler (1997).

Table 2 A chronology of taxing and spending interventions

Year	Events
1789	Committee on Ways and Means in House of Representatives begins handling both revenue and expenditure matters, a period of centralized budgeting begins
1791	Twenty year charter for the First Bank of the United States begins
1801	Beginning of Jefferson Administration (1801–1809)
1807	Embargo Act severely curtails trade and tax revenue
1811	Congress fails to renew charter of the First Bank of the United States
1812	War of 1812 begins (1812–1815)
1816	Twenty year charter for the Second Bank of the United States begins
1828	Tariff of 1828 known as the 'Tariff of Abominations'
1835	For the first and only time, all of the Federal government's interest bearing debt paid off; large increase in revenue from public land sales (1835–1836)
1836	Charter for the Second Bank of the United States expires
1837	Beginning of severe and lengthy recessionary period (1837–1843)
1843	Change in fiscal years, FY1843 only six months long
1846	Mexican War begins (May 1846–February 1848)
1861	Civil War begins (April)
1862	Tax Act of 1862 expands income tax and increases internal excise taxes on liquor and other items, country is divided into districts, more tax agents are added and the office of Commissioner of Internal Revenue is created (July)
1864	Tax Act of 1864
1865	Appropriations Committee carved out of Ways and Means Committee —dual committee system begins, one for taxes, the other for spending
1866	Huge decrease in military expenditures as army is reduced from 1 million to 50,000 men
1872	Income tax repealed completely after earlier reductions; tariff rates reduced as well
1885	Appropriations Committee stripped of much of its spending authority and given to numerous authorizing committees—period of decentralized budgeting begins
1890	McKinley Tariff; Pension Act
1897	Dingley Tariff (July)
1898	Spanish-American War (April–July)
1917	US enters WWI (April); War Revenue Act greatly expands income tax (October)
1921	Budget Reform Act restores power of Appropriations Committee—return to centralized budgeting; Revenue Act of 1921 raises exemptions and decreases income tax rates; Mellon becomes Secretary of the Treasury, a fiscal conservative
1932	Revenue Act increases a variety of taxes; Appropriations Committee begins to lose spending authority to subcommittees—return to decentralized budgeting
1933	Beginning of New Deal, Roosevelt inaugurated March 4
1941	US enters World War II (December)
1943	Current Tax Payment Act introduced employer withholding which made income tax a mass tax for the first time in US history
1946	Employment Act
1950	Korean War begins; Excess-Profits Tax enacted
1951	Individual income taxes raised for war finance; Korean War truce talks begin (February)
1953	Korean War truce signed (July), war taxes removed (December)
1964	Tonkin Gulf Resolution (August); major tax cut
1965	Major troop buildup in Vietnam begins (July)
1968	Tax surcharge enacted
1969	Surcharge removed; major tax bill; Vietnam troop withdrawal begins
1970	Buildup of entitlements and transfer payments begin
1975	Tax rebate
1976	Tax Reform Act
1978	Revenue Act
1980	Entitlements stabilize
1981	Reagan tax cuts
1982	Cut in budget share of nonmilitary purchases
1986	Tax reform

Table 3 Parsimonious characterizations of the conditional and marginal distributions

Regression equations		Summary statistics*									
Regression†	Equation‡	R^2	Standard error of regression	Sum of squared residuals	Normality $\chi^2(2)$	AR(1) AR(2)	F(1.,.) F(2.,.)	ARCH(1) F(1.,.)	Chow	F(.,.)	Nested UDL(1) F(.,.)
(i) XCP	$X_t = 0.0085 + 0.629X_{t-1} - 0.587X_{t-2}$	0.44	0.0020	0.000052	0.55	0.01	(1,12)	0.52	2.87	(3,10)	0.45
1791–1806	(0.0022) (0.225) (0.227)					0.22	(2,11)	(1.14)			(3,10)
(ii) XMP	$X_t = 0.0085 + 0.62X_{t-1} - 0.587X_{t-2}$	0.44	0.0020	0.000052	0.55	0.01	(1,12)	0.52	2.87	(3,10)	§
1791–1806	(0.00022)(0.225) (0.227)					0.22	(2,11)	(1,14)			—
(iii) RCP	$R_t = 0.541R_{t-1} + 1.014X_{t-2}$	0.57	0.0026	0.000095	0.66	0.03	(1,13)	0.19	0.62	(2,12)	0.48
1791–1806	(0.117) (0.248)					0.03	(2,12)	(1,14)			(4,10)
(iv) RMP	$R_t = 0.009 + 0.535R_{t-1}$	0.27	0.0034	0.000163	0.44	0.03	(1,13)	0.51	1.89	(2,12)	0.05
1791–1806	(0.005) (0.238)					0.07	(2,12)	(1,14)			(1,13)
(v) XCP	$X_t = 0.005 + 0.360R_t$	0.67	0.0019	0.000040	1.16	0.002	(1,10)	0.40	2.32	(2,9)	0.56
1816–28	(0.002) (0.077)					0.304	(2,9)	(1,11)	0.98	(2,9)	(4,7)
(vi) XMP	$X_t = 0.005 + 0.529X_{t-1}$	0.50	0.0023	0.000059	0.89	3.18	(1,10)	0.27	2.03	(2,9)	0.38
1816–28	(0.002) (0.159)					3.39	(2,9)	(1,11)	2.26	(2,9)	(1,10)
(vii) RCP	$R_t = 1.761X_t$	0.66	0.00041	0.000205	0.98	0.34	(1,11)	0.001	0.00	(1,11)	2.44
1816–28	(0.090)					0.82	(2,10)	(1,11)	0.07	(1,11)	(5,7)
(viii) RMP	$R_t = 0.030 - 0.384R_{t--2}$	0.20	0.0067	0.000490	3.51	3.81	(1,10)	1.33	2.11	(2,9)	0.52
1816–28‖	(0.005) (0.232)					2.51	(2,9)	(1,11)	1.22	(2,9)	(1,10)
(ix) XCP	Xt = 0.0140	0.00	0.0015	0.000025	0.87	0.34	(1,11)	2.75	0.65	(1,11)	0.90
1848–60	(0.0004)					0.16	(2.10)	(1,11)	0.92	(1,11)	(6,6)
(x) XMP	$X_t = 0.0140$	0.00	0.0015	0.000025	0.87	0.34	(1,11)	2.75	0.65	(1,11)	1.28
1848–60	(0.0004)					0.16	(2.10)	(1,11)	0.92	(1,11)	(2.10)
(xi) RCP	$R_t = 0.023 - 0.614X_{t-1}$	0.34	0.0021	0.000048	0.98	0.09	(1,10)	0.21	1.48	(2,9)	1.61
1848–60	(0.004) (0.2590)					0.20	(2.9)	(1,11)	1.56	(2,9)	(4.7)
(xii) RMP	$R_t = 0.0078 + 0.465R_{t-1}$	0.26	0.0022	0.000053	0.94	0.002	(1,10)	2.22	0.65	(1,11)	0.15
1848–60	(0.0034) (0.235)					0.005	(2,9)	(1,11)	1.47	(1,11)	(1,10)
(xiii) XCP	$X_t = 0.009 + 0.203R_{t-2}$	0.45	0.0013	0.000017	0.73	0.33	(1,10)	0.09	0.07	(2,9)	1.04
1872–84	(0.002) (0.068)					1.27	(2,9)	(1,11)	0.95	(2,9)	(4,7)
(xiv) XMP	$X_t = 0.0064 + 0.561X_{t-1}$	0.34	0.0014	0.000021	0.54	0.47	(1,10)	0.01	0.09	(2,9)	0.46
1872–84	(0.004) (0.235)					0.35	(2,9)	(1,11)	0.91	(2,9)	(1,10)

(xv) RCP	$R_t = 0.009 + 0.653X_{t-1}$	0.72	0.0018	0.000036	1.45	1.35	(1.10)	1.58	1.44	(2,9)	0.84
1872–84	(0.004) (0.122)					0.61	(2,9)	(1,11)	3.33	(2,9)	(4,7)
(xvi) RMP	$R_t = 0.009 + 0.653R_{t-1}$	0.72	0.0018	0.000036	1.45	1.35	(1.10)	1.58	1.44	(2,9)	3.36
1872–84	(0.004) (0.122)					0.61	(2,9)	(1,11)	3.33	(2,9)	(1,10)
(xvii) XCP¶	$X_t = 0.0076 + 0.294R_t + 0.256X_{t-1}$	0.72	0.0012	0.000015	0.78	0.01	(1,9)	0.34	0.57	(3,7)	0.16
1900–12	(0.0002) (0.218) (0.236)					1.91	(1,8)	(1,11)	0.41	(3,7)	(3,7)
(xviii)XMP	$X_t = 0.008 + 0.538X_{t-1}$	0.67	0.0013	0.000018	1.49	0.33	(1,10)	1.15	0.83	(2,9)	0.98
1900–12	(0.002) (0.113)					0.75	(2,9)	(1,11)	0.40	(2,9)	(1,10)
(xix) RCP	$R_t = 0.678R_{t-1} - 0.294R_{t-2}$										
1900–12	(0.165) (0.106)										
	$+0.313X_{t-1} + 0.303X_{t-2}$	0.98	0.00063	0.000003	0.67	0.02	(1,8)	0.04	0.13	(4,5)	0.53
	(0.087) (0.102)					0.38	(2,7)	(1,11)	0.43	(4,5)	(2,7)
(xx) RMP	$R_t = 1.330R_{t-1} - 0.364R_{t-2}$	0.91	0.0011	0.000014	0.96	0.38	(1,10)	1.76	1.56	(2,9)	2.99
1900–12	(0.170) (0.168)					0.56	(2,9)	(1,11)	0.38	(2,9)	(1,10)

Notes:

* The normality test reported is the C.M. Jarque and Anil K. Bera (1980) test for normal residuals, which is distributed $\chi^2_{[2]}$ under the null hypothesis of normality. AR(.) is a Lagrange-multiplier test for autocorrelated residuals. Suppose that the structural equation is $Y_t = a + bZ_t + e_t$, the LM statistic is based on the R^2 from the auxiliary regression: $e_t = a + bZ_t + g_1 e_{t-1} \ldots + g_1 e_{t-m}$, where e_t is the residual from the structural equation. The F-distribution equivalent is reported, which is distributed F(.,.) under the null hypothesis of no residual autocorrelation up to the order indicated by the degrees of freedom in the numerator. ARCH(1) is a Lagrange-multiplier test for first-order autoregressive conditional heteroscedasticity. The F-distribution equivalent is reported, which is distributed F(1,.) under the null hypothesis of no first-order autoregressive conditional heteroscedasticity. Explanatory power is a F test distributed as F(.,.) under the null hypothesis that all of the regressors, except the constant, are zero. Degrees of freedom are reported in parentheses. Chow = Chow test of first half of the sample period versus the second half. In cases where the sample contains an odd number of observations, two statistics are reported. For example, many samples contain 13 observations. The top number is the Chow test of the first six observations versus the last seven, while the bottom number is the first seven observations versus the last six. UDL(1) = F-test of exclusion restrictions versus corresponding unrestricted (general) distributed-lag model.

† XCP denotes the conditional, parsimonious specification for expenditures while XMP denotes the marginal, parsimonious specification for expenditures. Likewise, RCP denotes the conditional, parsimonious specification for revenues while RMP denotes the marginal parsimonious specification for revenues. Years correspond to the suspected tranquil periods.

‡ Beneath each coefficient estimate, the standard error is given in parentheses.

§ Section XMP 1791-1806 is identical to the corresponding unrestricted regression.

|| Every specification more parsimonious failed at least an ARCH(1), normality, and/or autocorrelation test.

Equation without X_{t-1} failed AR(2) test.

With these unrestricted regressions as a quality control, we seek a more parsimonious form containing only statistically significant regressors. These regressions are reported in Table 3. We require that the errors of the parsimonious form also show the desirable properties of no serial correlation, no conditional heteroscedasticity, and normality. With a few exceptions the regressions in Table 3 pass the tests. In some cases noted in the table, it was necessary to retain insignficant regressors in order to maintain white-noise errors. Beyond white noise errors, we require of each regression that it be a valid restriction of (i.e., that it not show significant loss of information compared to) the corresponding unrestricted regression. The test reported in the column 'Nested in UDL(1)' is an F-test of the relevant restrictions. Every regression passes at the 95% confidence level. As a further check against misspecification and mutually as a check on our assumption of the tranquility of our 'tranquil' periods, we perform a test of structural stability that checks the fit of the first half of the sample against the second half. This is reported in the column 'Chow'. Once again, all of the regressions pass this test.

We seek parsimonious forms in order to sharpen the power of the structural break tests. The essential feature of these regressions is that they convey the information contained in the unrestricted forms. It is important to recall that each unrestricted and restricted regression is a reduced form and, as such, it is inappropriate to place a structural interpretation on its parameters. We draw no inferences that require structural parameters. The question we want to ask is whether the characterizations of particular probability distributions are stable across specific sorts of interventions or regime changes. For that, we do not need structural parameters, but can live with coefficients that are complex functions of deeper parameters. Furthermore, because the estimates are for relatively short samples and because there need not be an underlying consistency in long-run budget decisions over the relatively short time-horizons of the tranquil periods, there is no reason that estimates might not imply unsustainable long-run relationships.

Our empirical work is bivariate. Partly that is because we have accounted for the role of some other variables in the construction of our regressors. Partly, it is because there are no other relevant data available. Of course, the real causal structure is unlikely to be bivariate. Nevertheless, we believe that the bivariate approach is unlikely to be misleading in this case. Omitted variables are a problem only if they lead us to identify spurious breaks because of the misspecification. The evidence presented below leads us to believe that, if other variables matter, they are either sufficiently random and unimportant or they matter in such a way that their omission does not induce us to identify spurious structural breaks. We conduct a battery of specification tests, including within-sample stability tests, to test for the likelihood that potentially omitted variables matter in an important way. As mentioned above, in almost all cases we cannot reject the hypotheses that the errors in the general unrestricted forms are white noise during the tranquil periods. Similarly, when the statistical tests are used to identify structural breaks outside of the tranquil periods, we carefully checked their plausibility against the historical record. We examined dozens of plots of recursive coefficients to buttress the formal tests of structural breaks and the historical record. Summaries of both these plots and the formal tests are presented in Table 4.[17] The congruence of both the historical and the statistical evidence convinces us that we have identified genuine structural breaks, and not spurious breaks owing to misspecification.

It is worth notice that in some cases the reported conditional regressions and marginal regressions are identical. This means, for example, that in an expenditures regression the receipts variables did very little to reduce the standard error of regression, so that they could be excluded without loss of information from the unrestricted form. That this sometimes happens in itself carries no information about causality, for it may occur simply because there just happened to be too little sampling variability in the particular sample to hand. Although the absence of a distinct conditional regression may hamper us somewhat in our causal inferences, we must nevertheless persevere and subject the marginal regression to the tests of structural stability.

4.4 Structural breaks in the non-tranquil periods

There are two distinct questions relating to relative stability. Does a structural break occur at all? And, given that it occurs, exactly when? A number of formal statistical tests have been developed to answer the first question. The second question remains less well understood and, at present, can be addressed only with informal methods.

Table 4 presents four tests of structural stability. Using recursive regression techniques (Harvey 1990, pp.52–6), each regression is projected backward and forward from the baseline tranquil period to the beginning (backward) or the end (forward) of the next tranquil period. Bracketing potential interventions with different tranquil periods and projecting forward from the earlier period and backward from the later period when testing for structural breaks serves several purposes. It permits us in some instances independently to confirm the timing of the breaks. When multiple breaks occur in the projection period it is sometimes easier to distinguish them when approaching the projection period from different directions. Finally, as noted above, sometimes the variation in the data will not permit us specify separate marginal and conditional regressions for some tranquil periods; so that we can sometimes fill in the information lost because the other adjacent tranquil period does permit distinct specifications.

The first stability test is the max-Chow test. The statistic is the maximum value of the set of Chow statistics computed for every possible break point in the sample. The statistic reported in the table is scaled by the 5% critical value (see Andrews, 1993), so that a value greater than unity indicates rejection.

---

[17] Eleven of these plots are reported in Hoover and Siegler (1997).

Table 4 Summary of structural break tests

Backward recursive coefficients	Backward max Chow statistic	Backward max Chow date	Backward fluctuation statistic	Backward fluctuation date	Backward sequential Chow date	Regression	Forward max Chow statistic	Forward max Chow date	Forward fluctuation statistic	Forward fluctuation date	Forward sequential Chow date	Forward recursive coefficients
						XCP-i (1791–1806)	NCM	NCM	NCM	NCM	NCM	NCM
						XMP-ii (1791–1806)	0.81	no break	3.30	1812	1812	1812
						RCP-iii (1791–1806)	1.26	1815	1.53	1813	1809	1814
						RMP-iv (1791–1806)	0.68	no break	0.80	no break	1809	no break
1812–16?	1.36	1812	1.12	1812	1815	XCP-v (1816–28)	1.45	1846	2.01	1836	1838	1837, 1847?
1812	0.65	no break	1.09	1821	1813	XMP-vi (1816–28)	0.86	no break	0.58	no break	no break	no break
1813	1.15	1809	0.68	no break	1815	RCP-vii (1816–28)	3.05	1836	2.14	1836	1838	1838
no break	0.98	no break	0.77	no break	no break	RMP-viii (1816–28)	2.36	1818	2.11	1838	no break	1838
NCM	NCM	NCM	NCM	NCM	NCM	XCP-ix (1848–60)	NCM	NCM	NCM	NCM	NCM	NCM
1848	1.01	1847	0.91	no break	1847	XMP-x (1848–60)	0.66	no break	0.68	no break	1862	1862
unstable	2.32	1834	2.48	1821	1847	RCP-xi (1848–60)	3.06	1865	2.86	1860	1861	1860–5
no break	1.86	1819	1.20	1837	1836	RMP-xii (1848–60)	1.32	1866	4.02	1866	1864	1865–7
1862, 1867	1.48	1862	1.96	1862	1867	XCP-xiii (1872–84)	1.65	1888	1.91	1889	1891	1886–91
1865–7	1.57	1866	1.41	1866	1868	XMP-xiv (1872–84)	0.83	no break	1.33	1899	1899	1891, 1899?
NCM	NCM	NCM	NCM	NCM	NCM	RCP-xv (1872–84)	NCM	NCM	NCM	NCM	NCM	NCM
1862?	1.32	1867	1.58	1864	1870	RMP-xvi (1872–84)	0.89	no break	1.56	1890	no break	no break
1890–9	1.10	1891	5.66	1895	1899	XCP-xvii (1900–12)						
no break	0.83	no break	1.07	1889	1899	XMP-xviii (1900–12)						
1885–99	0.82	no break	4.36	1892	1899	RCP-xix (1900–12)						
unstable	0.56	no break	0.62	no break	1899	RMP-xx (1900–12)						

Notes: The notation NCM (no conditional model) indicates that the parsimonious conditional distribution is the same as parsimonious marginal distribution. The notation '?' indicates uncertain identification of a structural break. Forward structural break tests extend to the end of the next tranquil period, while backward structural break tests extend to the beginning of the previous tranquil period.

The second test is the fluctuation test of Ploberger *et al.* (1989) modified by Chu (1990) to use the Euclidean norm rather than the infinite norm. The test compares the coefficient estimates for each recursive regression to those for the whole sample. The reported statistic is the maximum value scaled by its 5% critical value.[18]

The third test is a sequential Chow test. In it, the regression for the baseline period is compared to a sequence of regressions corresponding to the sequence of recursive regressions across the projection period. The plots of these statistics (scaled by their 5% critical values to maintain comparability) are examined visually.

There is a formal distribution theory for the first two tests, but none for the sequential Chow test. The distribution theory for the max-Chow and the fluctuation tests is asymptotic and the critical values for the tests and their statistical power have been determined from Monte Carlo experimentation. In any case, the tests only address the first question: does a break occur? On the second question—exactly where?—we must appeal to less formal results. There is some evidence that when there is only one break in a sample that the max-Chow test is likely to peak at or near the break. We therefore report the date at which the maximum is reached. Similarly, the fluctuation test is likely to peak at or near a single break. Were one to hypothesize a break at a particular point and run a Chow test to determine whether the null of stability could be rejected with that break point, the sequential Chow test would provide the sequence of those tests. Thus, one might reasonably surmise that the first time that the Chow test violated its 5% critical value would indicate that a break had occurred.

Experimentation also suggests that the most useful method of dating structural breaks is to examine the plots of the recursively estimated coefficients against their ± 2 standard error bands. Although we examine these in every case, too many plots are generated to be reported here. We therefore only summarize the results of our visual inspections in Table 4.

Tests of structural breaks are known to suffer from loss of power in small samples, which implies a tendency to find stability when in fact it does not exist. We are constrained by the data to use short samples. There is no one acid test of the existence of structural breaks or, more importantly, of their timing. Yet establishing the timing is the critical feature of the inferential procedure that underlies this project. The best one can do is judiciously to consider the results of the different tests in conjunction with the historical narrative to identify those times when the weight of evidence from different sources points to a structural break. While a rational scepticism is always in order, this appears to be the only way forward given the constraints we face, and the mutually reinforcing nature of the evidence gives us

[18] The fluctuation test with a Euclidean norm, which measures the variability of all the coefficients in a regression, is recommended only when there are five or fewer regressors. In the few instances in which there are more than five regressors, we substitute the original fluctuation test with an infinite norm, which measures only the variability of that coefficient which changes the most at each iteration of the recursive regression. When it is possible to run both tests, they almost always agree in any case.

reasonable confidence in our success. The detailed assessment of each individual case and the consilience of a variety of historical and econometric evidence are what convinces us of the correct identification of the breaks; nevertheless, this assessment is a tedious process and, in the interest of saving space and the reader's patience, we omit a detailed discussion of each case. (The interested reader should consult Hoover and Siegler, 1997, for the omitted details.) The outcome of our assessments is reported in Table 5 along with the events drawn from Table 2 that seem most nearly to correspond to the dates of the identified breaks. The evidence for some of the breaks is relatively weak. These are marked with a '?'.

5. The causal structure of fiscal policy in US history

Using Table 5 we can give a tentative assessment of the likely causal order between taxes and spending based on the different non-tranquil periods of US fiscal history. Table 1 is the key to interpreting Table 5. Table 1 suggests that one should look for distinctive patterns of three among the breaks: the situation in which both the marginal and conditional distribution on either the side of taxes or the side of expenditures breaks, while only one of the equations on the other sides breaks. If the event which corresponds to the break is on the same side as the paired conditional and marginal breaks, then we conclude that the statistics and the history

Table 5 Summary of structural breaks

Regressions				
Expenditures		Receipts		
Conditional	Marginal	Conditional	Marginal	Events
		1809(?)	1809(?)	Embargo Act (1807)
1812	1812			War of 1812 begins
		1813		
1815(?)	1815(?)	1815(?)		War of 1812 ends
1836–8		1836–8	1836–8	Debt paid off (1835); 2nd Bank of US end; deep recession; unusually large public land sales
1847	1847	1847		Mexican-American War
	1861–65			Civil War begins
1862		1863	1864–67	Tax Acts of 1862 and 1864
	1865, 1867	1866		House Appropriations Committee formed from Ways and Means; demobilization
			1870/71(?)	Income tax repealed
1886–91	1891/1892(?)	1891/1892(?)		Appropriations process reform (1885), McKinley tariff (1890)
1899	1899	1899	1899	Spanish American War; Dingley Tariff

Note: Notation '(?)' indicates a doubtful identification of a break.

are mutually reinforcing, and we are justified in concluding that the break corresponds to a genuine intervention and in assigning the break to that side. The remaining break allows us to discriminate the direction of causation as described in Table 1 and Section 3 above. Table 6 presents the results of our judgements of the implications of the breaks reported in Table 5. To summarize Table 6, in three of four non-tranquil periods there is evidence that taxes cause spending. The fourth period is relatively short and there is reason to believe, as we note below, that the implied change of causal direction occurred in the middle of the period. Furthermore, in order to explain the pattern of breaks, all of the tranquil periods must also be associated with taxes causing spending. Clearly, the pattern of taxes causing spending is the dominant causal pattern of US fiscal policy.

As with the assessment of structural stability, the reader is referred to Hoover and Siegler (1997) for detailed of the statistical evidence in relationship to US economic history, most of which is omitted here to save space. Some details must, however, be presented if the reader is to make sense of the mapping between Tables 5 and 6, which is not completely mechanical in every case.

In the period 1807–15, there are breaks at 1812 and 1813. To form a characteristic triad of breaks, we must associate them with the single intervention of the onset of the War of 1812. Given that the data are annual, this is reasonable, even though they are localized in different years.

The 1829–47 period is anomalous. The econometric and historical evidence is consistent with a short-lived reversal in the causal order, so that spending causes taxes. There are two main interventions in this period. The first is indicated by breaks in the receipts regressions and the conditional expenditures regression in the 1836–38 period. Without reference to the historical narrative, this is the configuration of a process in which spending causes taxes faced with a shock to the tax process. The four years 1835–38 were an extraordinary time. With the paying off of the entire national debt for the first and only time in history in 1835 (Stabile and Cantor, 1991, p.41) and the expiration of the charter of the Second Bank of the United States in 1836, fiscal institutions had reached a watershed. While the fiscal system is likely to have been shocked by these institutional changes, it is not completely clear how they affected the underlying tax and spending processes. An extraordinary period of government land sales and a severe depression (1837–43) are likely to have affected the revenue process, lending some weight to the statistical identification of a shock to the tax process.

Table 6 Summary of results

Non-tranquil Period	Causal direction
1807–15	taxes cause spending
1829–47	spending causes taxes
1861–71	taxes cause spending
1885–99	taxes cause spending

The second intervention in 1847, associated with the Mexican-American War, is indicated by the breaks in both expenditures regressions and the conditional receipts regression. This is a pattern indicative of an intervention in the expenditures process when the causal order runs from taxes to spending. A possible way to understand the effect of eliminating the national debt is suggested by Milton Friedman's view that deficits and the political necessity of debt repayment inhibit government expenditure.[19] Taxes cause spending so long as the debt constraint binds politically, but spending causes taxes when the constraint is removed. The Mexican-American War, with the revival of debt finance, restored the political constraint.

The analysis of the Civil War and Reconstruction periods (1861–71) requires an extended discussion as there are some nice issues regarding the timing of the breaks that are essential to a sensible interpretation. The investigation is somewhat hampered by the lack of a distinct conditional regression for receipts in the tranquil period before the Civil War. The breaks in the marginal expenditures regression clearly mark out the beginning and the end of the war—suggesting strongly that they are picking up the huge change in the level and composition of government expenditure associated with mobilizing and demobilizing a large army and navy. The military grew from 28,000 men in 1860, the last year of peace, to 1.1 million men in 1865 and then rapidly shrunk to 77,000 men in 1866, the year following surrender; by 1870, it had fallen further and hovered around 40,000 men until the outbreak of the Spanish-American War in 1898 (*Historical Statistics of the United States*, Series Y 904–916, and Studenski and Krooss, 1952, pp.163–4).

The breaks in the marginal receipts regression in the period 1864–67 come well into the war and continue after it. This is consistent with the historical record. Congress was slow to enact tax legislation because of the widespread belief that the South would be defeated rapidly (Studenski and Krooss, 1952, p.137; and Stabile and Cantor, 1991, p.52). The 1862 Tax Act was passed over a year after the beginning of the war. The tax act established a wide range of excise taxes and the nation's first income tax, which by 1866 generated 25% of Federal receipts. As with other tax acts, there were inevitable delays in its implementation. Studenski and Krooss (1952, p.150) observe that '... immediate collections under the 1862 Act were extremely disappointing ... [I]n the meantime, there was much confusion and evasion. Instead of $100 million, as expected, the internal-revenue collections in the fiscal year 1863 amounted to only $39 million. To increase revenue, Congress passed new tax laws in 1863, 1864, and 1865, those of 1864 being the most important' (see Studentski and Krooss, 1952, chs 13 and 14, for a much more complete discussion of Civil War finance). The income tax was reduced after the war and finally repealed in 1872, which may correspond to the weak evidence of a break dated at 1870/71 in the marginal receipts process.

[19] Friedman argued this case in a speech to the Commonwealth Club of San Francisco (7 August 1992).

Despite the fact that the Civil War presented major disruptions to both expenditures and revenue process, the timing of the breaks presents the possibility of a tentative interpretation. There appears to be an expenditure intervention in 1861–62. If we associate this with the break in the conditional receipts process in 1863 and assume that the later breaks in the marginal receipts process are more properly associated with the later major changes in tax structure, then there appears to be the classic pattern of an expenditures intervention in a causal order running from taxes to spending. If we regard the breaks between 1864 and 1867 in the conditional and marginal receipts process as related to the Tax Acts of 1862 and, especially, of 1864, as seems consistent with the narrative history, then the fact that there is evidence of a break in the marginal expenditures process but not in the conditional expenditures process in 1865–67, presents the pattern of a receipts intervention with a causal order again running from taxes to spending. The evidence of two different interventions within the Civil War, then, both point to the same conclusion: taxes cause spending in this period. The fact that several interventions occur in both receipts and expenditure processes in a very short run of years makes the evidence for the Civil War somewhat murky. The identification of causal direction here gains some cogency from the fact that it appears to be consistent with the causal structure identified in the immediately adjacent periods.

Every regression also indicates a break at 1899. This is consistent with two separate interventions—one on the tax and one on the spending side. These can be associated with the Spanish-American War and its aftermath. This is a common pattern in wartime. Such a situation does not indicate necessarily that causal order has changed, only that the data are such that no discrimination is possible.

A similar issue arises with respect to the twentieth century. Hoover and Siegler (1997) extended the investigation from the end of the 1900–12 tranquil period to 1990. Part of this work overlapped with that of Hoover and Sheffrin (1992). We present none of these results in this version, because it turns out that there are simply too many interventions, too close together in the period 1913 through World War II to uncover any useful discrimination. The structural break tests identify key dates in fiscal history, but they are unable to locate instances in which some regressions remain stable while others break, which is essential for inferring causal direction. The history of this period reveals that tax and spending interventions often came close together, so this is a unfortunate but hardly surprising. The period after World War II is covered in Hoover and Sheffrin (1992) and our results are consistent with theirs.

6. Some lessons

Hoover and Sheffrin (1992) found that the causal structure of US fiscal policy shifted in the relatively short time since the Second World War from taxes causing spending to causal independence between taxes and spending. The current study was motivated partly by the wish to determine whether such a shift in causal structure, which appeared to be the result of a fundamental change in the nature of the budgetary process, was unique to the 1970s and, if it was not, whether past shifts fit into a similar pattern. The evidence suggests that while taxes causing spending is the dominant causal pattern in US fiscal history, there is at least one episode of a reversal of this causal order in the nineteenth century.

It is important to recall the logic of the methodology of causal inference employed here. Interventions are revealing of causal order because they induce changes in the parameters of the underlying relationships. They do not necessarily alter the causal order when they induce these changes. It is possible that causal order can change in a time without any interventions. Thus, the dating of an intervention does not in itself tell us when the causal order changed. For that we must appeal to other information, particularly to the historical narrative. The dominant pattern of the nineteenth century is the same as the pattern after World War II up to about 1970: taxes cause spending. The intervention of 1836 to 1838 is then an extraordinary instance of a reversal of the dominant causal order. There was evidence that the reversal occurred at about the same time as the intervention. The new causal order was, however, short-lived. Again the causal order shifted by 1847 with the revival of debt finance associated with the Mexican-American War; and, given the history of the time, it is not unreasonable to assume that it shifted about 1847. The 20th century provides little good quality discriminating evidence.

Causal order in US fiscal policy is in practice, then, at least partially malleable. As we suggested in Section 1, this malleability is not surprising in the case of the fiscal process. What is perhaps surprising, given the possibility of fundamental change in causal structure, is the evidence that the pattern of taxes causing spending is dominant through most of US fiscal history.

One lesson of this history is that the dominant causal order does not seem to vary in any systematic way with dominant political party or even with budgetary structure. Cogan (1993) has argued that there are significant differences in budget behavior between periods of centralized budgeting (1799–1885 and 1922–31), which tend to run smaller deficits, and periods of decentralized budgeting (1886–1921 and 1932–present), which run larger deficits. A glance at Fig. 1 suggests that the wars dominate the history of deficits, at least up to 1954. Furthermore, the causal ordering of taxes and spending seems largely to be stable across Cogan's regimes. Hoover and Sheffrin (1992) identify the shift in the 1970s to causal independence between taxes and spending with a major change in Congressional budgetary procedures, but similar changes in the nineteenth century did not apparently induce similar shifts in causal order in every case. The 1885 reform appears to show up as an intervention in the causal process, but not as a causal reversal; while the 1836–38 episode appears to be merely a temporary shift in the causal order.

A second lesson is for economists. As we noted in Section 1 there are alternative theories of tax and spending behavior that implicitly define causal orders. It is not uncommon for economists to take one of them—say Barro's tax-smoothing model—and apply it to the whole of US history or, at the least, to some longish part of it. This could be right only if the causal order in history happened to agree with the order implicit in the model. The shifts of causal order uncovered in this paper suggest that *a priori* theoretical models are unlikely to perform well for any time chosen arbitrarily. This is professionally flattering for the economic historian; for it suggests that in reaching an effective understanding of US fiscal history, he will have a leg up on the ahistorical macroeconomist whose principal interest in the past is to find additional observations on which to fit his regression.[20]

References

Andrews, D.W.K. (1993). 'Tests for parameter instability and structural change with unknown change point', *Econometrica*, **61**, 821–56.

Barro, R.J. (1979). 'On the determination of public debt', *Journal of Political Economy*, **87**, 940–71.

Basmann, R.L. (1965). 'A note on the statistical testability of "explicit causal chains" against the class of "interdependent" models', *Journal of the American Statistical Association*, **60**, 1080–93.

Basmann, R.L. (1988). 'Causality tests and observationally equivalent representations of econometric models', *Journal of Econometrics*, **39**(Annals), 69–101.

Berry, T.Sr. (1988). *Production and Population Since 1789, Revised GNP Series in Constant Dollars*, Bostwick Paper No. 6, The Bostwick Press, Richmond, VA.

Calomiris, C.W. and Hanes, C. (1995). 'Historical macroeconomics and American macroeconomic history', in Kevin D. Hoover (ed.), *Macroeconometrics: Developments, Tensions and Prospects*, Kluwer, Boston, 351–416.

Cogan, J.F. (1993). 'Federal Budget', in D. R. Henderson (ed.), *The Fortune Encyclopedia of Economics*, Warner Books, New York, NY, 243–53.

Cooley, T.F. and LeRoy, S.F. (1985). 'Atheoretical macroeconometrics: A critique', *Journal of Monetary Economics*, **16**, 283–308.

Chu, C.J. (1990). *The Econometrics of Structural Change*, unpublished Ph.D dissertation, Department of Economics, University of California, San Diego.

Engle, R.E., Hendry, D.F. and Richard, J.-F. (1983). 'Exogeneity', *Econometrica*, **51**, 277–304.

Gilbert, C.L. (1986) 'Professor Hendry's econometric methodology', *Oxford Bulletin of Economics and Statistics*, **48**, 283–307.

Granger, C.W.J. (1980). 'Testing for causality: A personal viewpoint', *Journal of Economic Dynamics and Control*, **2**, 329–52.

Harvey, A. (1990). *The econometric analysis of time series*, 2nd edition, MIT Press, Cambridge, MA.

Hendry, D.F. (1983) 'Econometric modelling: The consumption function in retrospect', *Scottish Journal of Political Economy*, **30**, 193–220.

Hendry, D.F. (1987). 'Econometric methodology: A personal viewpoint', in Truman Bewley (ed.), *Advances in Econometrics*, Vol. 2. Cambridge University Press, Cambridge, 29–48.

[20] See Calomiris and Hanes (1995) for a general discussion of the importance of historical understanding in macroeconometrics sympathetic to our conclusion.

Hendry, D.F. and Richard, J.-F. (1982). 'On the formulation of dynamic models in empirical econometrics', *Journal of Econometrics*, **20** (Annals), 3–33.

Hoover, K.D. (1988). *The New Classical Macroeconomics: A Sceptical Inquiry*, Basil Blackwell, Oxford.

Hoover, K.D. (1990). 'The logic of causal inference: econometrics and the conditional analysis of causation', *Economics and Philosophy*, **6**, 207–34.

Hoover, K.D. (1991). 'The causal direction between money and prices: An alternative approach', *Journal of Monetary Economics*, **27**, 381–423.

Hoover, K.D. (2001). *Causality in Macroeconomics*. Cambridge University Press, Cambridge.

Hoover, K.D. and Sheffrin, S.M. (1992). 'Causation, spending and taxes: Sand in the sandbox or tax collector for the welfare state?' *American Economic Review*, **82**, 225–48.

Hoover, K.D. and Siegler, M.V. (1997). 'Two centuries of taxes and spending: A causal investigation of the federal budget process, 1791–1990', Working Paper No. 97–30, Economics Department, University of California, Davis.

Jacobs, R.L., Leamer, E.E. and Ward, M.P. (1979). 'Difficulties with testing for causation', *Economic Inquiry*, **17**, 401–13.

Lindert, P.H. (1994). 'The rise of social spending, 1880–1930', *Explorations in Economic History*, **31**, 1–37.

McAleer, M., Pagan, A.R. and Volker, P.A. (1983). 'What will take the con out of econometrics', *American Economic Review*, **75**, 293–307.

Mizon, G.E. (1995). 'Progressive modelling of macroeconomic time series: The LSE methodology', in K. D. Hoover (ed.), *Macroeconometrics: Developments, Tensions and Prospects*, Kluwer, Boston, MA, 107–70.

Moore, G.H. and Zarnowitz, V. (1986). 'Appendix A: The development and role of the National Bureau of Economic Research's business cycle chronologies', in R. J. Gordon (ed.), *The American Business Cycle: Continuity and Change*, University of Chicago Press, Chicago, IL, 735–79.

Neftçti, S. and Sargent, T.J. (1978) 'A little bit of evidence on the natural rate hypothesis from the US', *Journal of Monetary Economics*, **4**, 315–20.

Pagan, A. (1987). 'Three econometric methodologies: A critical appraisal', *Journal of Economic Surveys*, **1**, 3–24.

Phillips, P.C.B. (1988). 'Reflections on Econometric Methodology', *Economic Record*, **64**, 334–59.

Ploberger, W., Krämer, W. and Kontrus, K. (1989). 'A New Test for Structural Stability in the Linear Regression Model', *Journal of Econometrics*, **40**, 307–18.

Simon, H.A. (1953). 'Causal ordering and identifiability', in *Models of Man*, Wiley, New York, NY, 1957, 1–36.

Simon, H.A. and Iwasaki, Y. (1988). 'Causal ordering, comparative statics, and near decomposability', *Journal of Econometrics*, **39**, 149–73.

Stabile, D.R. and Cantor, J. (1991). *The Public Debt of the United States: An Historical Perspective, 1775–1990*, Praeger, New York, NY.

Studenski, P. and Krooss, H. (1952). *Financial History of the United States*. McGraw-Hill, New York, NY.

Trehan, B. and Walsh, C.E (1988). 'Common trends, the government budget constraint, and revenue smoothing', *Journal of Economic Dynamics and Control*, **12**, 425–44.

US Bureau of the Budget (1992). *Historical Tables, Budget of the United States Government, Fiscal Year 1992*, Government Printing Office, Washington, DC.

US Bureau of the Census (1975). *Historical Statistics of the United States, Colonial Times to 1970*, Government Printing Office, Washington, DC.

US Bureau of the Census (1992). *Statistical Abstract of the United States*, Government Printing Office, Washington, DC.

US Bureau of Economic Analysis (1992). *Survey of Current Business*, **72**.

Wildavsky, A. (1988). *The New Politics of the Budgetary Process*, Scott, Foresman, Glenview, IL.

Appendix

Data sources and definitions

The raw data for the expenditure and revenue variables come from two sources: for the period 1789–1900, US Bureau of the Census (1975); for the 20th century, US Bureau of the Budget (1992). Expenditures are total federal outlays net of interest payments while revenues are represented by total federal government receipts. All variables are converted from nominal to real values using a constructed price index described below. Furthermore, in order to abstract from the effects of inflation and cyclical movements of GNP on receipts and expenditures, we assume that the levels of taxes and spending are set relative to potential output. Therefore, all variables are scaled by potential output in order to remove common trends. Detailed information on each variable is provided below.

EXN = Nominal Federal Government Expenditures

1789–1900: US Bureau of the Census (1975), Series Y457, Total Outlays of the Federal Government. Note that 1789 = 1790 = 1791. Since fiscal year (FY) 1843 was only six months long, the reported series was doubled for 1843 in order to annualize the estimates.

1901–90: US Bureau of the Budget (1992), Table 1.1, Total Outlays. Note that 1901–1933 data are the same as Series Y457 above.

INTN = Nominal Interest

1789–1970: US Bureau of the Census (1975), Series Y461, Interest on the Public Debt. Note that 1789 = 1790 = 1791. Since fiscal year (FY) 1843 was only six months long, the reported series was doubled for 1843 in order to annualize the estimates.

1971–90: US Bureau of the Budget (1992), Table 8.2, Interest on the Public Debt. Note that the 1962–1970 data are the same as Series Y461 above.

RCPN = Nominal Federal Government Receipts

1789–1900: US Bureau of the Census (1975), Series Y352, Total Federal Government Receipts—Administrative Budget. Note that 1789 = 1790 = 1791. Since fiscal year (FY) 1843 was only six months long, the reported series was doubled for 1843 in order to annualize the estimates.

1901–90: US Bureau of the Budget (1992), Table 1.1, Total Receipts. Note that 1901–1933 data are the same as Series Y352 above.

P = Consumer Price Index (1967 = 100)

1789–99: US Bureau of the Census (1975), Series E52, Wholesale Price Indexes (Warren and Pearson), All Commodities. (WPI)

1800–1970: US Bureau of the Census (1975), Series E135, Consumer Price Index, All Items (1967 = 100), (CPI67)

1971–90: US Bureau of the Census (1992), No. 738, Consumer Price Indexes, All Items, 1960–91, (1982–84 = 100), (CPI82)

The three price series are linked in the following manner:

To construct a CPI for 1789–99 from the WPI, multiply the ratio of the two series at 1800 to the WPI for the years 1789–99.

$(Series\ E135_{1800}/Series\ E52_{1800}) = (51/129)$

$$CPI_i = (51/129)WPI_i \qquad \mathrm{i} = 1790, \ldots, 1799$$

To convert 1971–90 to 1967 = 100:

$$CPI67_i = (CPI82_i/33.4)100 \qquad \mathrm{i} = 1971, \ldots, 1990$$

where $CPI82_{1967} = 33.4$ with 1982–1984 = 100.

$GNPN$ = Nominal Gross National Product

1789–1928: Berry (1988), Table 9.

1929–90: US Bureau of Economic Analysis (1992), Table 1.

$EX = (EXN/P)100$ = Real Federal Government Expenditures

$INT = (INTN/P)100$ = Real Interest

$RCP = (RCPN/P)100$ = Real Federal Government Receipts

$GNP = (GNPN/P)100$ = Real GNP

$PGNP$ = Potential Real GNP

We seek a measure of potential output analogous to that implicit in official measures of capacity utilization. There are severe limits to the available data, but we believe that the following offers a good enough approximation for the problem at hand. Real GNP (*GNP*) was first converted from levels to logarithms. Secondly, a log-linear interpolated series was created by linking real GNP between NBER business-cycle peak years.[21] In several instances, real GNP exceeded the linearly interpolated series between the peak years. In order to get closer to the concept of capacity utilization, the entire piece-wise interpolated series was shifted up to become an 'envelope curve' tangent only at the year (1818), the farthest point above the peak-to-peak series. The entire peak-to-peak series was shifted up by 0.156279

[21] The NBER dates of business-cycle peaks are: 1796, 1802, 1807, 1812, 1815, 1822, 1825, 1828, 1833, 1836, 1839, 1845, 1847, 1853, 1856, 1860, 1864, 1869, 1873, 1882, 1887, 1890, 1892, 1895, 1899, 1903, 1907, 1910, 1913, 1918, 1920, 1923, 1926, 1929, 1937, 1944, 1948, 1953, 1957, 1960, 1969, 1973, 1979, 1981, 1990. See Tables A.2, A.3, and A.5 in Moore and Zarnowitz (1986, Appendix A).

since actual GNP in 1818 exceeded the interpolated series by nearly 16%. Consequently, potential output equals actual output in 1818 only. Lastly, *PGNP* was converted back to levels.

$X = (EX\text{-}INT)/PGNP$ = Real Federal Government Expenditures Net of Interest Payments as a Share of Potential Real GNP

$R = RCP/PGNP$ = Real Federal Government Receipts as a Share of Potential Real GNP