ECONSTOR
Make Your Publications Visible.

A Service of
ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Martineau, Nicolas-Guillaume; W. Smith, Gregor

Working Paper

Identifying fiscal policy (in)effectiveness from the differential adoption of Keynesianism in the interwar period

Queen's Economics Department Working Paper, No. 1290

Provided in Cooperation with:
Queen's University, Department of Economics (QED)

Suggested Citation: Martineau, Nicolas-Guillaume; W. Smith, Gregor (2012) : Identifying fiscal policy (in)effectiveness from the differential adoption of Keynesianism in the interwar period, Queen's Economics Department Working Paper, No. 1290, Queen's University, Department of Economics, Kingston (Ontario)

This Version is available at:
https://hdl.handle.net/10419/67864

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Queen's Economics Department Working Paper No. 1290

Identifying Fiscal Policy (In)effectiveness from the Differential Adoption of Keynesianism in the Interwar Period

Nicolas-Guillaume Martineau
Universite de Sherbrooke

Gregor W. Smith
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

1-2012

Identifying Fiscal Policy (In)effectiveness from the Differential Adoption of Keynesianism in the Interwar Period

Nicolas-Guillaume Martineau and Gregor W. Smith†

February 2012

Abstract

Differences across countries or decades in the countercyclical stance of fiscal policy can help identify whether the growth in government spending affects output growth and so speeds recovery from a recession. We use the heterogeneity in the government-spending reaction functions across twenty countries in the interwar period to identify this effect. The main finding is that the growth of government spending did not have a significant effect on output growth, so that there is little evidence that this central aspect of fiscal policy played a stabilizing role from 1920 to 1939.

JEL classification: E32, E65, N10

Keywords: fiscal policy, business-cycle history, Great Depression, interwar economy

†Martineau: Département d'économique, Université de Sherbrooke; Nicolas-Guillaume.Martineau@USherbrooke.ca. Smith: Department of Economics, Queen's University, Canada; smithgw@econ.queensu.ca. We thank the Social Sciences and Humanities Research Council of Canada and the Bank of Canada research fellowship programme for support of this research. We thank Steve Ambler, Hafedh Bouakez, Patrick Coe, Lee Ohanian, and seminar participants at the Canadian Economics Association meetings, the Bank of Canada, the T2M conference, McMaster University, and the Federal Reserve Bank of Philadelphia for very helpful comments. The opinions are the authors' alone and are not those of the Bank of Canada.

1. Introduction

Did the international experience of the interwar period show that expansionary fiscal policy hastened recovery from the Great Depression? Answering this question is an ambitious task, but we try to contribute to the answer by studying a large (for the time) panel of twenty countries from 1920 to 1939. An advantage of this breadth and of studying this time period is that there is great heterogeneity both in the business cycles experienced by these countries and in their fiscal policies. For example, France followed a largely passive fiscal policy, while Japan followed a more activist one. This variation across countries in cycles and policies should help us identify the role of fiscal policy. But a disadvantage of such a panel is that many macroeconomic measures — including consumption and real wages for example — are not available. So this approach is a complement to studies of individual countries, like the US and UK, where such series are available.

The limitations of the data, in frequency and coverage, may prevent us from reaching a precise answer about the efficacy of fiscal policy, but it is still of interest to know whether that is the conclusion. Of course, the answer and its precision depend on an identification scheme. This paper adopts a new one: the main identifying assumption is that counter-cyclical fiscal policy could have worked in any country but was not tried to the same extent in every country. Identification relies on differences across countries (or over time) in fiscal reaction functions that capture the response of government spending to national income. We use these differences to estimate the effect of this government spending on the growth of income in turn.

For simplicity we use the term 'Keynesian' to refer to fiscal reaction functions with a relatively counter-cyclical component. This term is slightly ahistorical for the interwar period, because *The General Theory* of course was not published and popularized until 1936. But Keynes advocated public works as an 'impulse' to reduce unemployment beginning in 1924 in *The Nation*. Both *Can Lloyd George Do It?* (1929) and the *The Means to Prosperity* (1933) argued for the 'cumulative effect' of public works. Admittedly, this term also does not do justice to the contributions of Swedish economists. For example, Myrdal described counter-cyclical government spending in a report for the Unemployment Commission in 1934 entitled *The economic effects of fiscal policy*. Montgomery (1938) instead referred to counter-cyclical fiscal policy as *konjunkturpolitik*, the adjustment of economic policy to cyclical change. Jonung (1979) used another contemporary term: *krispolitik*, the crisis policy. The roots of this policy lay in arguments by economists such as Wicksell and Ohlin, though according to Jonung it also was influenced by Keynes's writings of the 1920s.

Of course, business cycles in output also were affected by other shocks besides fiscal

policy changes. One example is the decline in US output transmitted to these countries as a decline in export demand. A second example is the stance of monetary policy, and specifically the timing of leaving the gold standard. We control for these shocks in trying to isolate the effect of changes in government spending. In fact, the identification depends on such observable, exogenous shocks, because output and government spending are treated as endogenous to each economy.

The paper measures a correlation between the growth of government spending and the growth of real output, controlling both for other influences on output and for the endogeneity of government spending. The idea is that this partial correlation should be present if fiscal policy had a significant effect, whatever the mechanism by which this occurred. There is a significant role for US output growth, so it is not the case that no statistical relationship can be detected because of the limitations of the data. But there is little evidence that government spending growth affected output growth.

2. Data and Design

The design of the study largely stems from the availability of data. We study twenty economies for the period from 1920 to 1939: Argentina, Australia, Austria, Belgium, Brazil, Canada, Czechoslovakia, Denmark, Finland, France, Germany, Hungary, Italy, Japan, the Netherlands, Norway, Portugal, Spain, Sweden, and Switzerland. The UK and US are included as influences on business cycles within this group, though we do not try to identify fiscal policy efficacy using their own policies.

We restrict the investigation to the interwar period for several reasons. First, data for the period before 1920 are even scarcer and of course would raise the issue of how to model wartime spending during 1914–1918. Second, data for the period after 1939 would raise the same issue for 1939–1945. The focus is on whether there is evidence for the efficacy of fiscal policy solely for the interwar period, rather than for long spans or later episodes. One advantage of this approach is that some research suggests the impact of fiscal policy changes is greater when interest rates are low; in this panel they tend to be low throughout the interwar period for most countries.

The data are at annual frequency. We selected all countries for which we could find measures of national income (GDP or GNP) and government spending. For most countries the government spending measure comes from public accounts rather than national accounts and so includes some transfers. We also use measures of defence spending collected by the League of Nations. The appendix contains details of the sources.

The government spending series we adopt begin in 1924 for Austria, Germany, and Hungary. Periods of hyperinflation thus are omitted from this study. Though it would

certainly be interesting to assess the impact of fiscal policy changes and to study the measurement of real government spending during hyperinflations, we cannot pursue those questions given the available data.

Figure 1 shows real national income per capita for each country, scaled so that 1920=100. The great heterogeneity in growth rates is evident, as countries such as Portugal, Japan, and Finland grew, albeit with cycles, while Austria stagnated. Figure 2 shows that there also was heterogeneity in real government spending per capita, where the role of the state expanded enormously in Japan while government spending also grew significantly in Germany, Sweden, Finland, and Hungary, for example.

Figure 3 shows government spending as a percentage share of national output for the 20 countries we study during the interwar period. The figure shows that there is also much variation across countries in the time path of this ratio. Thus the ordering of paths for output in figure 1 is not merely duplicated in the ordering of paths for government spending in figure 2, a feature which should aid identification. Figure 3 shows that there were sharp increases in government spending as a share of GDP, for example in Italy in 1936 and in France, Finland, and Hungary in 1939. There also were some sharp decreases, such as those in Belgium and Czechoslovakia during the 1920s. Notice also that the overall levels of these shares are in some cases very high; one probably cannot argue that the government sector was too small to have affected the economy.

It is worth noting some macroeconomic variables that are *not* available except for very narrow panels of countries: consumption, hours, markups, average tax rates, real wages and employment, and country-weights in international trade. We also considered unemployment rates as an alternate business-cycle indicator, but found that they were available for many fewer country-year combinations than was true of output. Thus we are not able to isolate the mechanism by which fiscal policy may have affected the path of national income. For example, we do not know whether fiscal policy affected the labour market via sticky wages, a wealth effect on labour supply, or a signal of future monetary expansion. Again, we use the word 'Keynesian' simply to refer to the relative counter-cyclicality of government spending and not to the underlying mechanism by which it may have an effect.

Data are available on tax revenue and hence on budget deficits. They are not included directly in the statistical model for several reasons. First, a benchmark model of the effects of tax timing — Ricardian equivalence — holds that the deficit may have any correlation with output. This absence of a prediction contrasts with the case of government spending, where a range of macroeconomic models predict a positive impact on output. Second, we generally do not have data on tax rates and their variation over time and across countries,

which would be necessary for the careful study of the impact of this aspect of fiscal policy. So we focus in this paper on government spending, like much recent research on post–1945 data. However, we do examine whether the impact of government spending varies with the level of the government budget deficit.

Overall the design of the study is to include as many countries as possible, particularly given the low frequency of the available data. We then study the simultaneous determination of government spending and national income within this panel.

3. Identification

The main aim of this paper is to investigate whether the heterogeneity in fiscal policy during the Great Depression can help us identify the impact of fiscal policy on output. We next show how this can potentially provide identification.

Let t count years and i count countries. Let y_{it} denote the growth rate of real output and g_{it} denote the growth rate of real government spending. Let x_{it} denote a weakly exogenous variable, such as output growth in the US or the UK. We study mainly small and medium-sized economies and so take US and UK output growth as given. We also discuss alternative exogenous variables below. The i subscript allows for later discussion of the identity of this exogenous variable to vary across countries.

Then consider a statistical model of domestic output growth as depending on the growth rate of government spending and on the foreign indicator:

$$y_{it} = \delta_{yi} + \beta_i g_{it} + \omega_i x_{it} + \epsilon_{yit}. \tag{1}$$

The parameter β_i measures the impact of the growth in government spending on the growth of output while the parameter ω_i measures the impact of an exogenous, observed shock x_{it}. The shock to output growth, ϵ_{it}, captures country-specific events such as the Japanese earthquake of 1923, the Credit Anstalt failure in Austria in 1931, or the French strikes of 1936.

Suppose that the reaction function for fiscal policy is:

$$g_{it} = \delta_{gi} + \alpha_i y_{it} + \epsilon_{git}. \tag{2}$$

The parameter α_i measures the response of government spending to the domestic business cycle. A negative value, for example, describes a counter-cyclical policy.

The intercepts δ_{yi} and δ_{gi} can differ across countries and across variables. The parameter δ_{yi} may reflect the productivity or capital accumulation trend in country i. The

corresponding intercept for government spending, δ_{gi}, may differ from that value. This difference can then capture the secular growth of the state in Japan during the interwar period, for example. By allowing for such growth in the output share of government spending we do not falsely confuse that with cyclical policy.

Estimating the parameters requires statistical assumptions, so we begin with:

$$\text{Assumption 1}: \ \mathrm{E}_i\, \epsilon_{yit} = \mathrm{E}_i\, \epsilon_{git} = 0 \ \ \forall i. \tag{3}$$

where E_i denotes the expectation over t in country i. The error terms have mean zero in each country. The sample versions of these moment conditions identify the intercepts.

The second assumption is that the observed, exogenous shock is uncorrelated with the unobserved shocks to output growth and to government spending growth:

$$\text{Assumption 2}: \ \mathrm{E}_i\, \epsilon_{yit} x_{it} = \mathrm{E}_i\, \epsilon_{git} x_{it} = 0 \ \ \forall i. \tag{4}$$

This is an OLS-type assumption that the regressor is uncorrelated with the error terms. Notice, though, that Assumption 2 does not rule out world shocks during the interwar period. Output growth y_{it} in country i can still be correlated with output growth in the US or UK through the term $\omega_i x_{it}$. And the shocks ϵ_{yit} can be correlated across groups of the 20 countries, or there can be country-specific shocks.

In the system (1)-(2) there obviously would be simultaneity bias in OLS estimation of either equation. Before elaborating on a consistent estimator, we should comment on the exclusion restrictions in the system. First, x_{it} is excluded from the g-equation (2); foreign output growth affects domestic output growth but does not directly affect government-spending growth. The idea is that the former effect stems from trade or financial linkages or from a common component in productivity shocks across countries, whereas it is difficult to see a reason why domestic fiscal policy should respond to the foreign business cycle. Second, while x shifts the y-equation (1) there is no corresponding shift variable exclusively in the g-equation (2). It is challenging to measure the exogenous component of elections, for example, and it seems unlikely that they would not also affect output. For this same reason any correlation between the two shocks is admissible; we do not assume $\mathrm{E}_i\, \epsilon_{yit}\epsilon_{git} = 0$.

Assumption 1 identifies the intercepts. Assumption 2 provides two moment conditions, but there are three remaining parameters, β_i, ω_i, and α_i, which thus cannot all be identified. An informative way to see the details is to solve the system to give the reduced form:

$$\begin{aligned} y_{it} &= \frac{\delta_{yi} + \beta\delta_{gi}}{1-\beta_i\alpha_i} + \frac{\omega_i}{1-\beta_i\alpha_i} x_{it} + \frac{\beta_i}{1-\beta_i\alpha_i}\epsilon_{git} + \frac{1}{1-\beta_i\alpha_i}\epsilon_{yit} \\ g_{it} &= \frac{\delta_{gi} + \alpha_i\delta_{yi}}{1-\beta_i\alpha_i} + \alpha_i\frac{\omega_i}{1-\beta_i\alpha_i} x_{it} + \big(1 + \alpha_i\frac{\beta_i}{1-\beta_i\alpha_i}\big)\epsilon_{git} + \frac{\alpha_i}{1-\beta_i\alpha_i}\epsilon_{yit} \end{aligned} \tag{5}$$

The composite error terms play no role in our identification, given Assumption 2, so we focus on the slopes, which we denote b_{yi} and b_{gi}, with:

$$\begin{aligned} b_{yi} &= \frac{\omega_i}{1-\beta_i\alpha_i} \\ b_{gi} &= \frac{\alpha_i\omega_i}{1-\beta_i\alpha_i}. \end{aligned} \tag{6}$$

It is easy to see that $\alpha_i = b_{gi}/b_{yi}$ and thus to see the textbook result that only α_i is identified. The exclusion of x_{it} from the reaction function allows α_i to be identified, but one cannot measure β_i, the impact of fiscal policy on the output growth rate y_{it}.

The reduced-form equations (5) also show that the correlation between y and g does not tell one about the effectiveness of fiscal policy. A positive correlation could reflect either (a) $\beta_i = 0$ and $\alpha_i > 0$ as growth allowed an expansion of the role of the state, or (b) $\beta_i > 0$ and $\alpha_i < 0$, with large enough shocks ϵ_{git}, as in a more Keynesian view. Similarly, ranking countries by the variance of their output growth rates does not rank them by the α_i, the extent of counter-cyclicality in fiscal policy, because the output-growth variance depends on the shock variance too, which may differ across countries.

The added identification relies on the heterogeneity of the reaction function across countries or over time, combined with a common fiscal-policy impact β, as we now show, with:

$$\text{Assumption 3}: \ \beta_i = \beta; \ \omega_i = \omega; \ \alpha_i \neq \alpha_j \ \ \forall i. \tag{7}$$

The impact of changes in government spending or changes in foreign output growth on the domestic business cycle is the same across countries. But the reaction function parameter, α_i, differs across at least one pair of countries. This is the differential Keynesianism referred to in the title of the paper.

To see identification it is enough to consider two countries labelled j and k. Assumption 2 now provides 4 moment conditions and, from Assumption 3, there are 4 parameters, β, ω, α_j, and α_k so this necessary condition for identification is satisfied. Again, the reduced-form equations provide the details. The 2 slopes in the reduced-form system vary by country giving 4 reduced-form coefficients. They are:

$$\begin{aligned} b_{yj} &= \frac{\omega}{1-\beta\alpha_j} \\ b_{gj} &= \frac{\alpha_j\omega}{1-\beta\alpha_j}, \end{aligned} \tag{8}$$

with similar equations in country k. It is straightforward to show that:

$$\begin{aligned} \alpha_j &= \frac{b_{gj}}{b_{yj}} \\ \alpha_k &= \frac{b_{gk}}{b_{yk}} \end{aligned} \tag{9a}$$

It is almost as straightforward to show that:

$$\begin{aligned}\beta &= \frac{b_{yj} - b_{yk}}{b_{gj} - b_{gk}} \\ \omega &= \frac{b_{gj}b_{yk} - b_{yj}b_{gk}}{b_{gj} - b_{gk}},\end{aligned} \tag{9b}$$

so that the parameters are just identified.

To see how to estimate β under Assumptions 1-3, consider basing estimation on the moment conditions:

$$\mathrm{E}_i(y_{it} - \delta_{yi} - \beta g_{it} - \omega x_{it} | \iota, x_{it}) = 0, \tag{10}$$

where ι is a vector of ones. Assumption 1 and the instrument ι identifies δ_{yi}. Then suppose there are only two countries, j and k, and so conditions (10) give two further moment conditions. To show that those conditions identify β and ω, ignore the intercepts and write the conditions (10) as:

$$\begin{aligned}\mathrm{E}_j y_{jt} x_{jt} - \beta \mathrm{E}_j g_{jt} x_{jt} - \omega \mathrm{E}_j x_{jt}^2 &= 0 \\ \mathrm{E}_k y_{kt} x_{kt} - \beta \mathrm{E}_k g_{kt} x_{kt} - \omega \mathrm{E}_k x_{kt}^2 &= 0.\end{aligned} \tag{11}$$

These conditions combine to give:

$$\begin{aligned}\beta &= \frac{\mathrm{E}_j y_{jt} x_{jt} / \mathrm{E}_j x_{jt}^2 - \mathrm{E}_k y_{kt} x_{kt} / \mathrm{E}_k x_{kt}^2}{\mathrm{E}_j g_{jt} x_{jt} / \mathrm{E}_j x_{jt}^2 - \mathrm{E}_k g_{kt} x_{kt} / \mathrm{E}_k x_{kt}^2} \\ &= \frac{b_{yj} - b_{yk}}{b_{gj} - b_{gk}},\end{aligned} \tag{12}$$

which reproduces our finding (9b). This rewriting thus shows that the moment conditions (10) use all the information on β from the solved, reduced-form system: having as many moments (or countries, in this case) as parameters is necessary and sufficient for identifying β.

Figure 4 gives the intuition. The upward-sloping line shows the effect of government spending growth on output growth, with slope β common to each country. The two downward-sloping lines show reaction functions, one for country j and one for country k. A single, observable shock x_t shifts the effectiveness curve in the same way for all countries. That shift traces out the reaction functions and so identifies their slopes α_j and α_k. But connecting the dots after such a shock, in other words pooling data from the two countries, also identifies the slope β, as shown.

Notice that we can think of the system as using a dummy variable which changes α and so rotates the reaction function. But the implied, additional variable in the reaction

function is the product of that dummy variable and y_{it}; that product is correlated with ϵ_{yt} and so is not a valid instrument. That means that country-specific dummy variables (or their products with output growth) cannot be used to identify β the way an exogenous shift in the reaction function would do.

A reduced form with slopes that vary across countries also could result from a structure in which the stance of fiscal policy α was the same across countries but its impact or efficacy, β_i varied across countries. But interpreting the reduced form as we have done is consistent with the way economists generally have tried to draw lessons from the macroeconomic experience of the Great Depression. For example, some economists and historians implicitly argue that counter-cyclical policy (a) was tried in some countries (like Sweden) but not others and (b) could have worked in the latter countries. We try to study the implications of this perspective.

It also is important to note that our approach does not require that governments pursued public works counter-cyclically with stabilization in mind. Any component of government spending whose cyclicality differed across countries can provide identification. For example, increased defence spending in Japan after 1931, in Germany after 1933, or in Italy associated with the invasion of Abyssinia in 1936 may provide identifying information. To take another example, Japan's fiscal expansion in 1931 while Takahashi was finance minister contrasts with the fiscal retrenchment in Germany under Chancellor Brüning at the same time, a difference that again may aid identification. However, the key feature is that these differences must be in part in the systematic part of government spending growth, linked to output growth by the parameter α_i; they cannot simply be differences in shocks ϵ_{git}.

To briefly formalize the identification requirements, suppose that there are H elements in x_{it} and I countries. Denote by n the cardinality (*i.e.* number of distinct elements) of a vector of parameters. Identification requires:

$$n(\beta_i) + n(\omega_i) \leq I \times H. \tag{13}$$

This is simply the usual method-of-moment requirement that the number of moments $(I \times H)$ in conditions (10) must be greater than or equal to the number of parameters. For example, when $I = 2$ and $H = 1$ one can identify β and ω. Thus whenever $I > 2$ one can either relax some restrictions on β_i and ω_i or gain efficiency and test the over-identifying restrictions. We explore some of these combinations of restrictions below. Not surprisingly, there is a plausibility vs. efficiency tradeoff involved in these restrictions, especially when there is so little data. But our main goal is not to see how many different β_i we can identify,

for no parameter is likely to be precisely estimated with so few time-series observations, but rather it is to estimate a common β across countries using these restrictions.

The second requirement is:

$$n(\alpha_i) \geq 2. \tag{14}$$

At least two distinct values of α are necessary for identification. As we add countries so that $I \geq 2$, further, distinct values of α further aid identification. (We do not show that formally because it is a straightforward extension of the earlier algebra.) We provide evidence below on this condition.

So far we imagine that the differences in fiscal-reaction parameters across countries apply throughout the interwar period; some countries were more Keynesian than others. But it also seems plausible that fiscal policy became more counter-cyclical over time, as Keynesian ideas spread. This change is not directly observable, unlike some breaks in monetary policy such as departure from the gold standard. But one can test for such a break within this statistical model, and exactly the same algebra just reviewed for two countries applies in comparing two time periods within a single country. Suppose that s and t index the 1920s and 1930s respectively and that we replace Assumption 3 with:

$$\text{Assumption } 3' : \ \alpha_{is} \neq \alpha_{it}. \tag{15}$$

then Assumptions 1, 2, and $3'$ allow identification country-by-country. In this case, country-specific values β_i and ω_i that are constant over time may be more plausible than ones that also are constant over countries as in Assumption 3. Alternately, we can identify parameters with a break in α_i in each country (or even the same break) combined with a common β and ω. Again these cross-country restrictions will add to statistical efficiency if they hold.

In sum, if the fiscal reaction varied significantly across countries or over time — a subject of independent interest — then there is scope for estimating the impact of government-spending growth on output growth. Identification comes from differences in the systematic component of fiscal policy, not from trying to isolate properties of exogenous changes in government spending.

The system (1)–(2) is in growth rates but involves no lags, a specification we choose in order to conserve degrees of freedom first because of the short span of annual data. A second reason for this choice is that a dynamic panel-data model with country-specific fixed effects would require further instruments for consistency, which again may be challenging. A third reason is that there are generally some missing observations in a historical panel like this one. For example, we do not have government spending data during the Spanish

civil war or during some of the central European hyperinflations of the early 1920s. A model of growth rates, as opposed to one with higher-order dynamics, gives stationarity but minimizes the loss of evidence from these missing data. Finally, we also inspect the residuals for evidence of autocorrelation and find no such significant pattern. We thus leave exploring added dynamics for further research with larger data sets. In that case, one also could apply the identification scheme to innovations instead of to growth rates, based on differences across countries or over time in how a shock to government spending is correlated with a shock to output.

The next section compares our method to those in recent research on fiscal policy's macroeconomic effects. It also gives references to previous research on fiscal policy specifically in the interwar period. Then section 5 reports the estimation results.

4. Research Context

This section very briefly outlines research on contemporary fiscal policy. One aim is to explain why we do not use either (a) a DSGE model or (b) a structural VAR model with unrestricted lags, which are the standard approaches. A second aim is to give the reader some benchmarks to which to compare our empirical findings below.

4.1 Fiscal Policy in Recessions

Isolating the effects of changes in fiscal policy can be challenging even with contemporary, quarterly data. Even for the US, where a long span of data is available, there is ongoing debate about whether conclusions about the effect of fiscal policy depend on specific, wartime episodes. Auerbach, Gale, and Harris (2010), Hall (2009), and Ramey (2011) review methods and findings from VAR methods and DSGE models as well as from microeconomic data.

DSGE modelling is challenging for the interwar period because we lack data on measures such as hours, real wages, markups, tax rates, and consumption reliably for the panel of countries. Such measures are needed to reliably calibrate the models and to measure shocks. For example, Cole and Ohanian (2011) find one can construct Solow residuals only for five countries in our panel (plus the UK and US) because of the lack of employment data.

In structural VAR models, identification of fiscal-policy shocks can come from (a) timing (*e.g.* the restriction that discretionary fiscal policy does not respond to the cycle within a quarter), (b) sign restrictions, or (c) the narrative approach (including using military spending). One challenge in applying SVARS to the interwar period is posed by the shortage of time-series observations because the data are annual. That also may make

timing restrictions less plausible. As Blanchard and Perotti (2002) note, it seems much less likely that government spending does not react to output within a year than within a quarter. That is even more true for a broad definition of government spending that includes some transfers.

Hall (2009) reviews multiplier findings from SVAR studies for the US. He shows that it is difficult to be precise about the impacts because there was not much variation in government spending except during World War II and the Korean War. Barro and Redlick (2009) study a long time series for the US to exploit this sort of variation. In a similar vein, we exploit the variation across countries during the interwar period. Their equation for output is similar to ours – output growth explained in part by growth in government purchases – but they also can include measures of expected growth in government spending and of tax rate changes; we do not have those series for these 20 countries. Barro and Redlick also argue that non-defense spending likely responds to the state of the economy; we use the variation in this response across countries to give identification.

A broad theme of much empirical research on the macroeconomic effects of government spending is that the impact of policy changes may be a variable that depends on such features as (a) the composition of spending (*e.g.* on infrastructure or transfers), (b) the financing (*e.g.* the timing of distorting taxes or the extent to which government spending signals future monetization), (c) the expected persistence of the policy. Unfortunately, most of these issues also cannot easily be studied for the interwar period because of data limitations. But this research also suggests that the impact of policy changes depends on (d) the state of the business cycle and (e) the presence of a liquidity trap. For example, Auerbach and Gorodnichenko (2010) find for the US that government spending multipliers are much larger in US recessions than in expansions. Parker (2011) also argues that one would expect the effect of a shock to government spending to be larger in a recession and also larger in a liquidity trap. Our study focuses on a period that includes the Great Depression, when many countries of course experienced both very high unemployment and very low interest rates. Finally, in the cross-country dimension, Ilzetski, Mendoza, and Végh (2010) argue that the efficacy of fiscal policy has recently depended on (f) the exchange-rate regime. We also control for cross-country differences in monetary (exchange-rate) policy.

4.2 Fiscal Policy in the Interwar Period

The effects of fiscal policy during the interwar period have been studied most extensively for the US and the UK. Crafts and Fearon (2010) survey research on the 1930s economy, mainly on those two countries. The upper panel of figure 5 shows the levels of real output per capita for the UK (in black) and US (in gray) for the interwar period,

scaled so that they begin at 100 in 1920. The differential experiences of these two countries during the Great Depression are well known, and reflected in the figure. By comparison with the business cycle in the UK, the depression in the US was (a) much more severe and (b) with a later trough (in 1933 rather than 1931 for the UK). The vertical scale is the same as in figure 1, which thus highlights the even greater heterogeneity of the experiences of the 20 countries we study.

Fishback (2010) provides a comprehensive review of US monetary and fiscal policy in the 1930s and outlines applications of DSGE models and VAR methods. Among recent studies, Eggertsson (2008) argues that expected future policy changed in 1933, in the form of a change in inflation expectations. He suggests that one of the triggers for the change in expectations was the large increase in government spending in 1933. Overall for the interwar period, though, Fishback reaches the same conclusion as Romer (1992) or Brown (1956, p 863) that

> fiscal policy, then, seems to have been an unsuccessful recovery device in the 'thirties not because it did not work, but because it was not tried.

Middleton (2010) describes fiscal policy in the UK in the 1930s and reaches a similar conclusion. However, some recent research for the US argues that increases in government spending did significantly affect the path of output in the very late 1930s and early 1940s. Gordon and Krenn (2010) conclude from a VAR model that government spending shocks played a large role in ending the US Great Depression after 1939 but before 1942. For the 1941–1945 period, McGrattan and Ohanian (2010) use a neoclassical DSGE model to measure the effects (for example in factor markets) of large fiscal shocks and conclude they had significant effects on output.

The lower panel of figure 5 shows government spending as a share of output in the UK (in black) and US (in gray). For the UK that share was very stable until 1938. For the US, the share rose gradually during the early 1930s, then levelled off, with a well-known dip in 1937 as some New Deal initiatives were reversed. The fiscal expansions in Italy or Japan were much more dramatic than that in the US. Our approach is based on the possibility that the extent to which fiscal policy was tried differed across countries, as figure 3 suggests. Of course, variation across countries in the government spending share of output does not automatically imply variation in the activist response to output, but we measure that latter variation explicitly in the next section.

Formal, statistical analysis of the role of fiscal policy in individual countries (other than the US and UK) during the interwar period usually is infeasible because of the absence of data at greater than annual frequency. But there are some noteworthy exceptions to this rule. Ritschl (2002) provides a very thorough review of research on fiscal policy in

Germany in the 1930s and an assessment and update of the macroeconomic data, which are more detailed than for most countries in the panel we study. He estimates a VAR model and concludes that government budget deficits had no significant effect on the path of German output; rearmament simply crowded out private spending. Cha (2003) estimates a VAR using a range of monthly data for Japan from 1929 to 1936 and concludes that fiscal shocks did play a large role in its relatively rapid recovery from the Depression.

Histories of the Depression in small, open economies sometimes comment on the stance and possible impact of fiscal policy. Feinstein, Temin, and Toniolo (2008, pp 194-199) provide a guide to these histories. Examples include the works by Safarian (1970) for Canada or Montgomery (1938) and Myrdal (1939) for Sweden. Safarian's (1970) book on the Great Depression in Canada developed a theme that has been echoed by other research: he attributed much of the downturn and recovery to events in the US and UK. Jonung (1979) and Grytten (2006) describe the post-war consensus view that Swedish fiscal policy was effective in reducing the scale of the Depression in Sweden. But they also concur that economists have been correct to question this consensus since the 1970s. Jonung noted that formal econometric analysis of the sources of recovery was lacking.

Even some contemporary observers doubted whether the *krispolitik* had a significant effect on recovery. Myrdal (1939) argued that counter-cyclical policy in Sweden was largely a failure in part "because this policy was carried out only half-heartedly" (p 183) with public works programs "of much smaller scope than would have been desirable" (p 184) and in part because of the "adverse reaction of business confidence, which has too often restricted or even possibly reversed its stimulating effects" (p 187). The share of government spending in GDP in Sweden rose from 8.2% in 1930 to 12.4% in 1933, though some other countries experienced sharper fiscal expansions.

Montgomery (1938), in what was surely one of the earliest books on recovery from the Depression, attributed Sweden's recovery to floating the exchange rate (leaving the gold standard with the UK) and to foreign recovery particularly in the UK. On the effect of fiscal policy changes, however, Montgomery (1938, p 67) also noted:

> This question can best be answered by the use of the comparative method. For the purposes of an enquiry of this kind we find the ideal type when those countries which are to be compared agree in other respects but differ in that particular point the significance of which we intend to investigate. We can hardly expect to come across such ideal types in the world of reality ...

But we may come across types as close to ideal as possible in this group of countries, precisely because Sweden, for example, followed a proto-Keynesian policy recommended by an array of eminent economists while some other countries did not.

Two recent studies do use international panels to study the interwar period, based on the idea that the heterogeneity across countries in the scale of the Depression should be informative. Cole and Ohanian (2011) use a DSGE model with shocks to productivity and the money supply to study 18 countries for the 1929–1936 period. They find a role for productivity shocks but not for money shocks. In our panel and time period we unfortunately cannot measure productivity shocks for even a majority of countries. However, we do allow for country-specific fixed effects in growth rates that may capture some of the cross-country, real-side differences detected by Cole and Ohanian.

Almunia, Bénétrix, Eichengreen, O'Rourke, and Rua (2010) study an international panel of 27 countries from 1925–1939. They estimate an SVAR model with four variables: defence spending, real GDP, government revenues, and a central bank discount rate. Tests show that one lag is significant in these annual data, and dummy variables by year and country also are included. Their identification then uses a Cholesky decomposition by ordering the variables, with defence spending not reacting to other shocks within the year. They find large effects of government spending shocks, perhaps due to the low interest-rate environment, but small monetary-policy effects. They thus conclude that fiscal policy may well have been effective had it been implemented more widely. Almunia *et al* also consider other orderings, removing fixed-effects because of possible inconsistency in a dynamic panel, and estimation with total government spending. With total spending they also estimate a single-equation model where output growth depends on country and year effects as well as the growth in government spending and the change in the central bank's discount rate, instrumented with defence spending and gold standard participation. Again they find significant effects of changes in government spending.

As in their second approach, we study total government spending. We model growth rates but do not estimate higher-order dynamics given the short span of annual data, though we do test for residual autocorrelation. Our potential contributions are based on using the cross-country variation in fiscal reaction to assess efficacy (rather than measuring shocks) and on conditioning on business cycles in the US and UK to provide instruments. Our findings from this identification end up being different from those of Almunia *et al*: there is little evidence that government spending growth had a significant effect on output growth overall for the 1920–1939 period.

5. Identification from Differences in Fiscal Reaction across Countries

We first present results from the statistical model outlined in section 3. There, we base identification on cross-country differences in the fiscal-reaction parameter, α, and treat output growth in the US and UK as exogenous. Then we consider controlling for

exchange-rate changes, allowing β to vary across sub-samples, and estimating the impact of the growth in defence spending.

5.1 Conditioning on US and UK Growth

Our benchmark, statistical model conditions on output growth in the US, y_{ut}, and output growth in the UK, y_{kt}, which thus compose x_{it} for all countries. Including these as exogenous variables is designed to capture the effect of the business cycle in the US and UK on the demand for exports from the countries in the panel. It may also capture an effect of the US and UK business cycles that operated through financial linkages. Initially we estimate parameters ω_u and ω_k on US and UK growth that are common across the 20 countries, but we then also allow these to vary over groups of countries and over time. For example, they may be lower during the relatively autarkic 1930s than during the 1920s.

These foreign variables also can capture the presence of a common, world shock. A good example is the banking crisis of 1931 that occurred in Austria, Germany, the US, and other countries. We lack a good financial crisis indicator for each country, but these variables may reflect that common shock. This feature of the statistical model also may capture a common technology shock that we cannot observe directly. For example, Betts, Bordo, and Redish (1996) find evidence of a large, real-side shock, common to both the US and Canada in the 1930s. They argue that it was a common shock, perhaps rather than transmission through export demand, a conclusion based in part on study of the terms of trade.

Define $z_{it} = \{\iota \;\; x_{it}\}$ as the instrument set. Then the estimating equations for GMM are the sample versions of:

$$\mathrm{E}_i\big[(y_{it} - \delta_{yi} - \beta g_{it} - \omega x_{it}) \cdot z_{it}\big] = 0, \tag{16}$$

for $t = 1921, ..., 1939$ and $i = 1, ..., 20$ countries. Any correlation between ϵ_{yit} and ϵ_{git} is possible, so there are no cross-equation restrictions between the y-equations (16) and the g-equations:

$$\mathrm{E}_i\big[(g_{it} - \delta_{gi} - \alpha_i y_{it}) \cdot z_{it}\big] = 0 \tag{17}$$

which thus can be estimated separately with no loss of efficiency. Estimation is by continuously-updated GMM and so standard errors are robust to heteroskedasticity and autocorrelation.

Table 1 shows results. In the first row $x_{it} = y_{ut}$ and both β and ω are restricted to be the same across all 20 countries. The estimated impact of government-spending growth is positive, but statistically insignificant at conventional levels of significance. But this is

not because nothing is significant; the impact of US output growth on domestic output growth, measured by $\hat{\omega}_u$, is positive and significant. Naturally, a significant x-variable is necessary for identifying β, and table 1 shows that y_{ut} qualifies.

Table 1 also shows the J-test statistic of the over-identifying restrictions, labelled J_y, along with its p-value. In this case the test readily rejects the hypothesis that β and ω are the same across all 20 countries. The next two rows therefore relax the restrictions on the parameters on y_{ut}. Weights that differ across countries could reflect geography, comparative advantage, or other causes of differential trade links to the US. They also can capture differences in exposure to the financial crisis of the early 1930s.

In the second row the responses to US output growth, labelled ω_{un}, differ across $N = 4$ groups of countries, with the groups sorted by continent: Argentina and Brazil; Australia and Japan; Canada; and then the 15 European countries. The third row has finer distinctions, with 9 groups of countries: Argentina and Brazil; Australia and Japan; Canada; Finland, Norway, and Sweden; Belgium, Denmark and the Netherlands; France and Switzerland; Italy; Spain and Portugal; and Austria, Germany, Hungary, and Czechoslovakia. In each of these cases $\hat{\beta}$ remains insignificant at the 5% level. The p-value associated with the J-test statistic also rises, and the restrictions now are not rejected at conventional levels of significance, which suggests that this disaggregation is sufficient.

The next three rows repeat the analysis but with UK output growth y_{kt} also included as an exogenous variable. It too at first has a common coefficient across countries, to conserve degrees of freedom. Unlike US output growth y_{ut}, this variable is statistically insignificant. Again, as we sort the responses to both US and UK output growth, labelled ω_{un} and ω_{kn} respectively, the coefficient on government spending growth, $\hat{\beta}$, remains statistically significant, while the p-value for the J-test statistic rises so that the restrictions are no longer rejected at conventional levels of significance. Overall, then, the statistical model passes this diagnostic test, with a significant role for US output growth but not for domestic government spending growth.

One possible explanation for the lack of precision in estimating β is that, while there was some variation in the degree of counter-cyclical policy across countries, our instruments are only weakly correlated with g_{it}. In this case, inference about β may be affected by the syndrome of weak identification, which can lead to imprecise estimates and also to hypothesis tests with incorrect size. Recall from section 3 that identification relies at least one difference across countries in the fiscal-reaction parameters, α_i. To check on the evidence, table 2 presents estimates of the 'first-stage' moment conditions (17), with $x_{it} = y_{ut}$. We first test whether $\alpha_i = \alpha$ for all countries. The first row reports the common estimate $\hat{\alpha}$, the test statistic, and its p-value. The p-value is 0.52, so we cannot reject the

hypothesis of a common fiscal-reaction parameter across the 20 countries. We also found relatively large p-values for this test statistic within the 9 groups of countries for which we tried to estimate separate values of β in table 1. Again this raises the possibility that identification is weak.

Imprecision in $\hat{\alpha}_i$ for individual countries may lead to low power in this test, so we also allow α to take on a country-specific value α_i while all other countries have a common value α_i^c. We then test whether the two values are equal. Table 2 reports the results for countries with p-values less than 0.10. In this case, we can reject the null hypothesis of a common value, at this level of significance, for four countries. According to the estimates, Finland and Portugal had government-spending growth rates which were significantly more pro-cyclical than those of the remaining countries, while the Netherlands and Spain had more counter-cyclical paths. These differences suggest that there is enough cross-country variation in α_i to identify β. But further research on tools for inference robust to weak identification (based on cross-equation restrictions) may be appropriate.

We also inspected the 20 residual series visually for evidence of autocorrelation. There is no such evidence, though perhaps no test will be very informative with this span of annual data. The findings so far of course also depend on both the conditioning variables x_{it} and on the identifying assumptions. We next consider additional x-variables and an alternative measure of government spending. We then turn to alternative identifying assumptions in section 6.

5.2 Conditioning on Exchange-Rate Changes

We next control for an indicator of monetary policy. Choudri and Kochin (1980) reported on the insulating properties of floating exchange rates for European countries such as Spain during the Great Depression. Eichengreen and Sachs (1985), Eichengreen (1992), and Bernanke (1995) have argued persuasively that the event of leaving the gold standard can be treated as exogenous and that it hastened recovery in a broad cross-section of countries. Eichengreen (2004) provides a comprehensive review and interpretation of research on the international Great Depression that also supports this idea. While their studies focus on the 1930s, there also was great heterogeneity in monetary policy in the 1920s. For example, the depth of the UK depression in the early 1920s often has been attributed to the Bank of England's restrictive monetary policy aimed at restoring the gold standard at pre-war parity. Given this research, we next control for exchange-rate changes, viewed as an indicator of monetary policy for the interwar period, by including them as an x-variable when we try to measure the impact of changes in fiscal policy.

We first include the growth rate of the local-currency price of gold. Thus large values reflect loose monetary policy. This variable may capture the well-known effect that

reflation of the domestic economy was the means by which monetary policy accelerated recovery. Section 2 noted that government spending series for Austria, Germany, and Hungary begin in 1924, after their hyperinflations. We thus omitted these extreme observations in testing for this effect and used only the depreciations of the new currencies (the shilling, reichsmark, and pengö respectively) in these countries. Figure 6 shows the time series in levels, with large, cumulative depreciations in Portugal, Brazil, France, Japan, and Spain.

We studied this rate of depreciation (a) as the sole regressor, with a common coefficient across countries, (b) alone but allowing for different coefficients across groups of countries, and (c) with US output growth also included. The rate of depreciation was statistically insignificant (at conventional levels) in each of these cases, as was the growth rate of government spending. With disaggregated effects, in case (b), the point estimates of the impact of depreciations are positive for France and Switzerland and negative for Australia and Japan, so there is some evidence of heterogeneity across groups of countries in this effect on output growth. But these responses also are insignificant, so that one cannot say that pooling conceals offsetting effects. And when US output growth is included it remains highly significant, essentially reproducing table 1.

How can one reconcile this evidence with the well-known statistical findings of Bernanke (1995)? First, we study 1921–1939 whereas he focused on the recovery from the Great Depression during 1931–1935. Second, he grouped countries according to whether they were on or off the gold standard, whereas we measure the rate of depreciation against gold for each country. He reported a significant effect of leaving the gold standard on a range of economic indicators. We study real GDP growth, whereas his indicators included manufacturing production and employment, and we also have data on slightly different groups of countries.

To examine the effect of these differences we also estimated the output equations (a) with the on/off gold indicator used by Bernanke (rather than the rate of depreciation), (b) for 1931–1935, and (c) with dummy variables for each year, as in his study. We found a positive correlation between being off gold during the early 1930s and output growth, but not a statistically significant one. The results were similar when we used the on/off gold indicator for the entire 1921–1939 time period.

Given Eichengreen's (1992) conclusions about exogeneity, being on or off the gold standard seems a better candidate for an exogenous variable than the scale of depreciation, which could well respond to business-cycle conditions. But this indicator too is statistically significant for output growth for the entire interwar period. Perhaps that is not surprising, for the 1920s of course included years in which a number of countries, including Argentina, Finland, France, Italy, and Sweden, were off the gold standard yet following monetary

contractions to try to restore their gold-standard parities. Pooling those episodes with the subsequent departures from the gold standard in the early 1930s leads to very little correlation between the exchange-rate regime and output growth.

A second, traditional way to measure exchange rates is relative to other currencies. This indicator measures relative reflation but also may provide information on a mechanism for recovery: depreciation that promoted net exports. We alternately included the rate of depreciation of each domestic currency against the US dollar or against sterling. Both these candidates for inclusion as exogenous variables were statistically insignificant also. Finding x-variables to be statistically significant is necessary for identification here, but we do not find that for exchange-rate depreciations. Omitting them thus does not explain the insignificant values of $\hat{\beta}$ found in table 1.

We conclude our discussion of controlling for depreciations with two important qualifications. First, finding a fool-proof indicator of the exchange-rate regime is not easy even in postwar data. Rose (2011) notes the lack of consensus on how to classify these regimes and reviews present-day classification methods and their impact. An obvious example of the pitfalls in such indicators arises for Austria, Czechoslovakia, Germany, and Hungary, which remained officially on the gold standard in the 1930s but applied a wide range of exchange controls. Their nominal exchange rates thus do not provide accurate measures of domestic reflation. Of course, exchange controls also were widespread in the early 1920s.

Second, the statistical insignificance of the exchange-rate indicators in this panel does *not* mean that adhering to, then leaving, the gold standard was not a central cause of the interwar business cycle. Our statistical model is compatible with the possibility that monetary policies in the US and the UK were important contributors to the business cycle in the world economy, as stressed by Feinstein, Temin, and Toniolo (2008). They report on the simultaneous monetary contractions in the US and Germany in 1931 for example. But we cannot isolate an *additional* role for country-specific exchange rate changes in these other economies. (In the case of fiscal policy, in contrast, few scholars have argued that government-spending shocks were important in the UK and US business cycles, so our finding that we cannot find a role for them in these 20 countries reinforces that conclusion.)

5.3 Sorting the Effect of Government Spending Growth

So far we have pooled all observations to allow identification and estimate the average effect of government-spending growth as precisely as possible. But this pooling may obscure some heterogeneity in β, the parameter that measures the impact of government-spending growth on output growth. We next explore possible heterogeneity along several dimensions in which we have reliable data.

Imagine that β is not a constant but depends on some variable v_{it}. Unfortunately, we cannot usually add an interaction term $v_{it}g_{it}$ to the y-equation, for v_{it} sometimes will be endogenous. Thus, we simply sort the observations into groups based on the value of several criterion variables v_{it} to see if that reveals variation in $\hat{\beta}$. Of course there is a limit to this sorting; for example we cannot identify country-specific values.

The first candidate as a sorting criterion is the value of the government budget balance, as a percentage of national income. The idea of course is that the impact of a change in government spending might depend on the financing method, perhaps with loan-financed changes having a larger effect than tax-financed ones. Once again, the interwar period is an interesting time to study, because there was large variation in budget balances over time and countries, which should be informative. For example, Belgium, Czechoslovakia, France, and Italy in the early 1920s all ran budget deficits that often exceeded 10% of GDP. Other European countries, like Denmark and Switzerland, remained near budget balance throughout the interwar period. A number of countries, such as Canada, swung from surplus to deficit during the 1930s. Japan ran surpluses throughout the period.

We sort by observations (not countries) according to the value of the budget surplus or deficit. Then we re-estimate the y-equation within each group of observations. Table 3 contains the results. The upper two rows divide the observations into two groups, while the lower three rows divide them into three groups again ranked by budget balances. The x-variable is US output growth, with a common coefficient within each group of observations.

The point estimate $\hat{\beta}$ is positive in each group of observations, but it remains statistically insignificant at conventional levels of significance. When we divide the sample in half, the point estimate is larger for the sample with large budget deficits. But when we divide the sample in thirds the point estimate is largest for the sample with the largest surpluses. Meanwhile, US output growth remains statistically significant in each sub-sample. The evidence from the J-test is mixed, with p-values ranging from 0.02 to 0.69, depending on the sub-sample. Overall then, this first sorting does not alter the conclusions.

A second way to sort observations is according to the share of government spending in GDP. One might expect that the impact of a change in the growth rate of GDP would be smaller in countries or observations in which government spending is a relatively small proportion of national income. Again, there is considerable variation in this ratio in these historical data. Figure 3 shows the government spending share varying widely across countries and over time. When we sort observations by this ratio and estimate the y-equation on sub-samples (not shown), we again find a highly significant role for US output growth for each sub-sample, but generally no significant effect of government-spending

growth. The only exception occurs for the third of observations with the lowest government spending shares, where there is a significant *negative* estimate $\hat{\beta}$.

Our third sorting is by decade. Some modern theoretical research suggests that the impact of changes in government spending may depend on level of interest rates and specifically that β may be larger when the economy is in a liquidity trap. (Ramey (2011), however, reports that there little evidence of this pattern in US history.) For our historical panel, figure 7 shows short-term, market interest rates for a number of countries. Although there is variation across countries in the overall level of interest rates, there is a trend down from the 1920s to the 1930s for many countries, including France, Canada, Japan and Sweden, for example. We do not have complete interest-rate data to incorporate in the statistical model, but we test whether β took on a different value in the 1930s from that in the 1920s. The results (not shown) show that $\hat{\beta}$ is insignificantly different from zero for both decades. Thus, the finding of an insignificant $\hat{\beta}$ for the entire interwar period is not the result of incorrectly pooling the two decades.

Finally, we also sorted the observations according to the exchange-rate regime and specifically by each of (a) whether the gold standard was in place or not, (b) the rate of depreciation against gold, and (c) the rate of depreciation against the US dollar. This sorting is designed in part to detect the traditional, Mundell-Fleming view that fiscal policy is more effective under a fixed nominal exchange rate than under a floating one. However, once again none of these ways of sorting the observations gives a sub-sample with a statistically significant coefficient $\hat{\beta}$ on government spending growth. In fact, the point estimate is negative for low rates of nominal depreciation and positive for high rates of depreciation.

5.4 Defence Spending

We also can study whether the growth of defence spending influenced real output growth during the interwar period. This investigation should be informative, for two reasons. First, there are sharp spikes up in defence spending in a range of years and countries, such as Japan during 1932–1934, Italy during 1935–1936 and Australia, Canada, Portugal, and other countries in 1939, as well as spikes down in Argentina during 1937 and France during 1936. This variation should help identify the impact, if any, on GDP growth. Second, the impact parameter, β, may be larger for defence spending than for other categories of spending or transfers, for theory suggests it may be the most stimulative via its welfare-reducing effect on labour supply. Admittedly, focusing on defence spending does not answer the question of whether changes in government spending overall affected the path of output, but it may tell us about a large component of the public-sector budget and whether it did matter to overall growth or could have done so.

We draw on data on defence spending from the League of Nations yearbooks, which yield 313 annual observations. We adopt the same statistical model as for total government spending and so allow for defence spending to respond to the business cycle. The results are very similar to those for total spending, shown in table 1, and so are not shown separately. There is a significant role for US output growth, and there is evidence of differences in α_i, the fiscal-reaction parameter for defence spending, across countries and especially over time. But the impact coefficient $\hat{\beta}$ is statistically insignificant. Its point estimate sometimes is negative, depending on the set of cross-country restrictions. In addition, we also find no significant role for defence spending when we treat its growth rate as exogenous, and so estimate by ordinary least squares.

6. Over-Identification from Differences in Fiscal Reaction over Time

So far the fiscal-policy reactions we have considered differed across countries but not over time; we have not allowed for an adoption of Keynesian policies (or any other changes in their cyclicality) that took place within the interwar period. Recall that it also is possible to gain identification from time-variation in α, as in Assumption $3'$. To explore this possibility we required restrictions (16) to hold separately for the 1920s and for the 1930s for each country; for a total of 40 restrictions. If α_i varied across decades in several countries then allowing for this variation should aid identification of the fiscal-policy impact parameter β.

Table 4 contains the results. We use the statistic J_y to test the over-identifying restrictions. The only exogenous regressor is y_{ut}, the US output growth rate. The first row applies a common β and ω_u to each decade and country, finding the former to be small and statistically insignificant, while the latter is significant as in table 1. But the J_y test rejects the restriction that these parameters are constant over time and countries.

The second row then allows a different value of ω_u, labelled ω_{un}, for each of $N = 9$ groups of countries: Argentina and Brazil; Australia and Japan; Canada; Finland, Norway, and Sweden; Belgium, Denmark and the Netherlands; France and Switzerland; Italy; Spain and Portugal; and Austria, Germany, Hungary, and Czechoslovakia. Again the idea is that the impact of US output growth on domestic output growth may well have varied across regions. Now there are fewer restrictions, but the J_y test statistic still rejects them readily.

The third row allows ω_{un} to vary both across groups of countries and between decades. This feature allows for the possibility that the impact of the US business cycle fell during the 1930s due to the worldwide collapse in trade (in turn caused in part by tariff increases and other autarkic policies), for example. Now the J_y statistic is just at the 5% critical value. Some research suggests that this test may over-reject, so there is now relatively

little evidence against the restrictions. Thus we can highlight the finding on the lack of impact of government-spending changes: $\hat{\beta}$ is small and insignificantly different from zero.

We could relax over-identifying restrictions by allowing β to vary across countries too, but we do not do so for two reasons. First, there are good economic reasons to believe ω_{un}, the response to US output growth, might be larger for Canada, for example, than for Sweden or Hungary. We do not have such a reason in the case of β. Second, the aim of the study is to gain precision in estimating β by pooling information across countries. With the added restrictions in table 4 we could estimate β and ω separately for each country, but when we do so they are all statistically insignificant, an unsurprising result given the short time span and low frequency of the data. In any case, with 20 countries, one $\hat{\beta}_i$ could be statistically significant at the 5% level simply due to randomness.

Section 5.1 commented on the potential weak identification problem that would arise with limited variation in α_i across countries. Table 2 showed evidence of distinct values for several countries. The further estimation in table 4 relies on α differing between the 1920s and 1930s for at least one country. To provide some evidence on this variation, we estimate a common fiscal-reaction for the 1920s that differs from the reaction for the 1930s. We find a significant difference between the two coefficients. That difference is sufficient to identify β, but it also is of independent interest. Notably, when we estimate the reaction function (17) pooled across countries we find a large, significant *increase* in $\hat{\alpha}$ from the 1920s to the 1930s, so that on average government spending became less counter-cyclical. The estimate (with standard error in brackets) rises from -1.33 (0.55) in the 1920s to 0.49 (0.41) in the 1930s. The difference between these two values of α, is 1.83 (0.62), a shift which provides identifying information for β. Thus, there is information in the variation across decades in average fiscal reaction, but (a) the change is in the direction of *less* counter-cyclicality and (b) exploiting this change to help identify the effectiveness parameter β does not change the conclusion that it is insignificantly different from zero.

7. Conclusion

The title of this paper refers to counter-cyclical policy as Keynesian, for simplicity. But well before the publication of *The General Theory*, Swedish economists already recommended *konjunkturpolitik*, the adjustment of economic policy to cyclical change. This paper adopts an identification scheme that seems to capture the spirit of many informal assessments of the role of this type of fiscal policy in the interwar period. First, it was tried to a greater extent in some countries than in others (so the reaction-function parameter α_i varied over countries). Second, it could have been used more widely (so the impact parameter β was the same across countries).

A study of fiscal policy in a panel of many countries in the interwar period cannot rely on measures of hours or consumption that one would naturally study in assessing more recent fiscal policy. Instead, we look for a correlation between the growth of government spending and the growth of real output, controlling both for other influences on output and for the endogeneity of government spending. The idea is that this partial correlation should be present if fiscal policy had a significant effect, whatever the mechanism by which this occurred.

Identification relies on an exogenous variable, and we find that US output growth was statistically significant for this panel of economies. Thus, it is not the case that no relationship can be statistically detected given the data limitations. Next, the last rows of tables 1 and 4 show that the statistical model passes the J-test so that the over-identifying restrictions cannot be rejected. In this context, then, could fiscal policy have been effective? When we base identification on cross-country differences in reaction-functions (in the last row of table 1), the 95% confidence interval for β is (-0.14, 0.18). When we also use information from changes in fiscal reaction functions over time (in the last row of table 4) the 95% confidence interval for β is (-0.08,0.26). In each case the interval includes zero and the upper edge is less than the average impact of the US business cycle.

Non-correlation does not necessarily imply non-causation. It is possible that there is some other variable that affected output and that we have a downward-biased estimate of the fiscal-policy effect because of omitting that. But to account for the non-correlation here such a variable would need to have been both (a) correlated with output growth over time and (b) correlated with government-spending growth over time but with the opposite sign, systematically across countries. It seems more likely that the average impact of government-spending growth simply was small, so that it could not have had a large effect on the interwar business cycle.

Assessing the extent to which stabilization with government spending was attempted is a more challenging question. We generally cannot measure the fiscal-reaction coefficient α_i with much precision, precisely because the identification relies on allowing it to vary over countries and decades. But there is significant counter-cyclicality in some countries and pro-cyclicality in others, as well as an overall trend to less counter-cyclicality from the 1920s to the 1930s.

So far, the case that these policies mattered to output growth cannot be proved using this identification and annual macroeconomic data for twenty countries. Finding an insignificant effect of government spending growth on output growth for the interwar period is especially noteworthy because this period included many years with high unemployment, very low interest rates, and large changes in defence spending, a combination

of circumstances that recent research suggests should be particularly likely to foster a macroeconomic impact for changes in government spending. But the evidence for these countries better fits the Treasury view.

8. Appendix: Data Sources

8.1 Countries and Currencies

The twenty countries are Argentina, Australia, Austria, Belgium, Brazil, Canada, Czechoslovakia, Denmark, Finland, France, Germany, Hungary, Italy, Japan, Netherlands, Norway, Portugal, Spain, Sweden, and Switzerland. The United Kingdom and the United States are the reference countries.

Five countries changed currencies or had multiple currencies during the period: Argentina the peso papel and the peso oro; Austria the schilling and the crown; Belgium the franc and the belga; Germany the reichsmark and papiermark; and Hungary the pëngo and korona. We take these changes into account in constructing exchange rates.

8.2 Nominal Output

Annual, nominal output from 1920 to 1939 come from Michael Bordo's financial crises dataset `http://econweb.rutgers.edu/bordo/Financial_Crises_Database.zip`, which also gives original sources, except for data for Austria (1924–1937), Czechosolvakia (1920–1937), and Hungary (1925–1939), which come from Mitchell (1998). Output is measured as GDP for Argentina, Australia, Brazil, Czechoslovakia, Denmark, and Norway and as GNP for other countries.

8.3 Population

Annual population series are from Bordo (nd) except for data for Czechoslovakia from Mitchell (2003).

8.4 Government Expenditures

Estimates of current government spending on goods and services on a national accounts basis for the interwar period are scarce. Data are available from Liesner (1989) only for Australia, Canada, Germany, Italy, Japan, and Sweden, and these series do not all begin in 1920.

For the full set of countries the measure of government spending is broader and includes some transfers. Such total government expenditure data are obtained from Mitchell (1998, 2003). These data are for central governments. The following states were unitary or highly centralized: Belgium, Brazil (especially after 1930), Czechoslovakia, Denmark, France, Finland, Germany (especially after 1933), Hungary, Italy, Japan, the Netherlands, Norway, Portugal, Spain (until 1932), and Sweden. For more decentralized or federal states there is the possibility of changes in local government expenditure offsetting those at the central level. Such states were Argentina, Australia, Austria (though with no state tax revenue), Brazil (1920–30), Canada, Germany (1920–33), Spain (1932–1939), and Switzerland.

Due to the data not being expressed in calendar years for all countries, and due to differences in the reporting of data for fiscal years, two types of moving averages were computed to approximate calendar-year expenditures from fiscal-year expenditures, wherever applicable. For Australia, Denmark, Hungary, Italy, Norway, and Sweden the data was

reported for the fiscal year ending in the given calendar year (*e.g.* 1920 data actually corresponding to 1919–20 fiscal-year data), so the moving average for that year was calculated using the fiscal year that followed (*e.g.* the average of 1919–20 and 1920–21). In contrast, for Canada, Germany, and Japan the data was reported for the fiscal year ending in the next calendar year (*e.g.* 1920 data actually corresponding to 1920–21 fiscal-year data), so the moving average for that year was calculated using the fiscal year that preceded it. Finally, some countries (France, Portugal, and Sweden) changed fiscal years and so required weights that varied over time in order to approximate calendar-year data from fiscal-year data. The details are available from the authors.

8.5 Deflators

Real output and government expenditures are estimated by dividing the nominal series by cost-of-living indexes from Bordo except for data for Czechoslovakia, Hungary, and Spain which are from Mitchell (1998). For Austria and Germany, prices are in terms of gold after currency stabilizations in 1922 and 1923 respectively. To avoid hyperinflationary periods we include Austria from 1923 and Germany and Hungary from 1924.

8.6 Government Revenue

Government revenues are from Mitchell (1998, 2003). Exact definitions of revenues differ from country to country, with some total revenue figures being only for taxation. After being adjusted from a fiscal-year to a calendar-year basis (as described for expenditures above), they are used with total government spending to calculate nominal government budgetary balance in each year. The ratio to nominal GDP is then used as the budget balance share.

8.7 Exchange Rates

Exchange rates are measured first in national currency per USD, from the League of Nations (1926–1940). They are converted to gold values by multiplication by the USD value in terms of gold from Officer and Williamson (2010).

8.8 Unemployment Rates

Unemployment rates are from Mitchell (1998) or Maddison (2003), which present different measures. But there are no data for Argentina, Brazil, Czechoslovakia, Hungary, Portugal, or Spain, and those for France, Italy, and Japan are for limited years. The total number of country-year observations was 199, compared to the 369 observations on real output.

8.9 Defence Spending

Annual defence spending is from the League of Nations (1924–1940) yearbooks, then deflated by the national cost-of-living index.

8.10 Interest Rates

Interest rates are short-term commercial paper or Treasury bill rates, where available, for 13 countries, from Bordo or the League of Nations.

References

Almunia, Miguel, Agustín S. Bénétrix, Barry Eichengreen, Kevin H. O'Rourke, and Gisela Rua (2010) From Great Depression to Great Credit Crisis: Similarities, differences, and lessons. *Economic Policy* 25, 219–265.

Auerbach, Alan J., William G. Gale, and Benjamin H. Harris (2010) Activist fiscal policy *Journal of Economic Perspectives* 24:4, 141–164.

Auerbach, Alan and Yuri Gorodnichenko (2010) Measuring the output responses to fiscal policy, mimeo, Department of Economics, University of California Berkeley.

Barro, Robert J. and Charles J. Redlick (2009) Macroeconomic effects from government purchases and taxes. NBER working paper 15369.

Bernanke, Ben S. (1995) The macroeconomics of the Great Depression: A comparative approach. *Journal of Money, Credit and Banking* 1–28.

Betts, Caroline M., Michael D. Bordo, and Angela Redish (1996) A small open economy in depression: Lessons from Canada in the 1930s. *Canadian Journal of Economics* 29, 1–36.

Blanchard, Olivier and Roberto Perotti (2002) An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *Quarterly Journal of Economics* 117, 1329–1368.

Bordo, Michael D. (nd) Financial Crises Database. `http://econweb.rutgers.edu/bordo/Financial_Crises_Database.zip`

Brown, E. C. (1956) Fiscal policy in the thirties: A reappraisal. *American Economic Review*, 46, 857–879.

Cha, Myung Soo (2003) Did Takahashi Korekiyo rescue Japan from the Great Depression? *Journal of Economic History* 63, 127–144.

Choudhri, Ehsan U. and Levis A. Kochin (1980) The exchange rate and the international transmission of business cycle disturbances: Some evidence from the Great Depression. *Journal of Money, Credit and Banking* 12, 565–74,

Crafts, Nicholas and Peter Fearon (2010) Lessons from the 1930s Great Depression. *Oxford Review of Economic Policy* 26:3, 285-317.

Cole, Harold L. and Lee E. Ohanian (2011) Cartelization policies and the international Great Depression. mimeo, Department of Economics, UCLA.

Eggertsson, Gauti B. (2008) Great expectations and the end of the Depression. *American Economic Review* 98, 1476–1516.

Eichengreen, Barry (1992) *Golden Fetters: The Gold Standard and the Great Depression, 1919–1939.* New York: Oxford University Press.

Eichengreen, Barry (2004) Viewpoint: Understanding the Great Depression. *Canadian Journal of Economics* 37, 1–27.

Eichengreen, Barry and Jeffrey Sachs (1985) Exchange rates and economic recovery in the 1930s. *Journal of Economic History* 45, 925-46.

Feinstein, Charles H., Peter Temin, and Gianni Toniolo (2008) *The World Economy between the World Wars.* Oxford University Press.

Fishback, Price V. (2010) U.S. monetary and fiscal policy in the 1930s. *Oxford Review of Economic Policy* 26, 385–413.

Gordon, Robert J., and Robert Krenn (2010) The end of the Great Depression, 1939-41: Policy contributions and fiscal multipliers. NBER Working Paper No. 16380.

Grytten, Ola Honningdal (2006) Why was the Great Depression not so Great in the Nordic Countries? mimeo, Department of Economics, Norwegian School of Economics and Business Administration.

Hall, Robert E. (2009) By how much does GDP rise if the government buys more output? *Brookings Papers on Economic Acitivity* Fall, 183–231.

Ilzetski, Ethan, Enrique G. Mendoza, and Carlos Végh (2010) How large (small?) are fiscal multipliers? mimeo, Department of Economics, London School of Economics

Jonung, Lars (1979) Knut Wicksell's norm of price stabilization and Swedish monetary policy in the 1930s. *Journal of Monetary Economics* 5, 459–496.

League of Nations (1924–1940) *Armaments Year-book, 1924–1940.* Geneva: League of Nations.

Liesner, Thelma (1989). *One Hundred Years of Economic Statistics: United Kingdom, United States of America, Australia, Canada, France, Germany, Italy, Japan, Sweden.* 2nd ed. London: The Economist Publications.

Maddison, Angus (2003) *The World Economy: Historical Statistics.* Paris: OECD

McGrattan, Ellen R. and Lee E. Ohanian (2010) Does neoclassical theory account for the effects of big fiscal shocks? Evidence from World War II. *International Economic Review* 51, 509–532.

Middleton, Roger (2010) British monetary and fiscal policy in the 1930s. *Oxford Review of Economic Policy* 26:3, 414–441.

Mitchell, Brian R. (1998) *International Historical Statistics, 1750–1993.* 4th ed. London: Macmillan.

Mitchell, Brian R. (2003) *International Historical Statistics: Europe 1750–2000.* 5th ed. Basingstoke: Palgrave Macmillan.

Montgomery, Arthur (1938) *How Sweden overcame the depression, 1930–1933.* Stockholm: A. Bonniers boktryckeri.

Myrdal, Gunnar (1939) Fiscal policy in the business cycle. *American Economic Review (P)* 29, 183–193.

Officer, Lawrence H. and Samuel H. Williamson (2010) The price of gold, 1257 – present. *MeasuringWorth* `http://www.measuringworth.com/gold/`

Parker, Jonathan A. (2011) On measuring the effects of fiscal policy in recessions. *Journal of Economic Literature* 49, 703-718.

Ramey, Valerie A. (2011) Can government purchases stimulate the economy? *Journal of Economic Literature* 49, 673-685.

Ritschl, Albrecht (2002) Deficit spending in the Nazi recovery, 1933–1938: A critical reassessment. *Journal of the Japanese and International Economies* 16, 559-582.

Romer, Christina D. (1992) What ended the Great Depression? *Journal of Economic History* 52 (December): 757–84.

Rose, Andrew K. (2011) Exchange rate regimes in the modern era: Fixed, floating, and flaky. *Journal of Economic Literature* 49, 652–672.

Safarian, A.E. (1970) *The Canadian Economy in the Great Depression.* Toronto: McLelland and Stewart.

Table 1: Fiscal-Policy Impact 1921–1939

$$\mathrm{E}(y_{it} - \delta_{yi} - \beta g_{it} - \omega x_{it} | \iota, x_{it}) = 0$$
$$t = 1921, \ldots, 1939$$

ωx_{it}	$\hat{\beta}$ (se)	$\hat{\omega}_u$ (se)	$\hat{\omega}_k$ (se)	$J_y(df)$ (p)
$\omega_u y_{ut}$	0.05 (0.09)	0.29 (0.04)	—	40.8(18) (0.00)
$\omega_{un} y_{ut}$ $N = 4$	0.12 (0.09)	-0.28/0.46 ((0.10/0.15)	–	22.7(15) (0.09)
$\omega_{un} y_{ut}$ $N = 9$	0.02 (0.10)	-0.26/0.47 ((0.13/0.17)	–	12.2(10) (0.27)
$\omega_u y_{ut} + \omega_k y_{kt}$	0.03 (0.07)	0.27 (0.04)	0.06 (0.10)	59.5(37) (0.01)
$\omega_{un} y_{ut} + \omega_{kn} y_{kt}$ $N = 4$	0.07 (0.07)	-0.18/0.40 (0.11/0.16)	-0.63/0.36 (0.29/0.41)	39.1(31) (0.15)
$\omega_{un} y_{ut} + \omega_{kn} y_{kt}$ $N = 9$	0.02 (0.08)	-0.23/0.38 (0.16/0.18)	-0.59/0.46 (0.38/0.48)	22.8(21) (0.35)

Notes: There are $I \times T = 350$ observations; y is real output growth; g is real government spending growth; y_u and y_k are growth in the US and UK. Estimation is by continuously updated GMM. The J_y-statistic tests the overidentifyingrestrictions. For parameters that vary across N groups of countries, entries are the minium and maximum estimates, with their standard errors.

Table 2: Cross-Country Differences in Fiscal Reactions

$$\mathrm{E}(g_{it} - \delta_{gi} - \alpha_i y_{it} | \iota, y_{ut}) = 0$$

$$t = 1921, \ldots, 1939$$

Countries	$\hat{\alpha}_i^c$ (se)	$\hat{\alpha}_i$ (se)	$J_g(df)$ (p)
All	-0.07 (0.25)	–	17.9(19) (0.52)
Finland	-0.26 (0.27)	2.82 (1.08)	9.1(1) (0.00)
Netherlands	-0.04 (0.26)	-3.90 (2.54)	3.2(1) (0.07)
Portugal	-0.08 (0.28)	6.11 (2.73)	3.4(1) (0.07)
Spain	-0.07 (0.56)	-2.57 (3.09)	14.7(1) (0.00)

Notes: y_u is US output growth. $J_g(19)$ tests $\alpha_i = \alpha \ \forall i$ while $J_g(1)$ tests $\alpha_i = \alpha_i^c$, the common value for all countries except i.

Table 3: Sorting the Fiscal-Policy Impact by Budget Balance

$$\mathrm{E}(y_{it} - \delta_{yi} - \beta g_{it} - \omega_u y_{ut} | \iota, x_{it}) = 0$$
$$t = 1921, \ldots, 1939$$

Surplus Range	$\hat{\beta}$ (se)	$\hat{\omega}_u$ (se)	$J_y(df)$ (p)
-29% to -1.6%	0.12 (0.07)	0.36 (0.05)	28.2(15) (0.02)
-1.6% to 12.6%	0.04 (0.09)	0.16 (0.06)	14.2(15) (0.51)
-29% to -2.8%	0.08 (0.06)	0.32 (0.06)	19.2(14) (0.16)
-2.8% to -0.6%	0.05 (0.09)	0.35 (0.07)	28.4(15) (0.02)
-0.6% to 12.6%	0.10 (0.12)	0.16 (0.07)	9.12(12) (0.69)

Notes: There are $I \times T = 350$ observations; y is real output growth; g is real government growth; y_u is growth in the US. The budget balance is a share of GDP. Estimation is by continuously updated GMM. The J_y- statistic tests the overidentifying restrictions.

Table 4: Fiscal-Policy Impact 1921–1939

$$\mathrm{E}(y_{it} - \delta_{yi} - \beta g_{it} - \omega_u y_{ut} | \iota, y_{ut}) = 0$$

$$t = 1921, \ldots, 1929 \quad \text{and} \quad t = 1930, \ldots, 1939$$

ω_u	$\hat{\beta}$ (se)	$\hat{\omega}_u$ (se)	$J_y(df)$ (p)
ω_u	0.04 (0.06)	0.28 (0.04)	78.2(38) (0.00)
ω_{un} $N = 9$	0.03 (0.06)	–	55.4(30) (0.00)
ω_{un}: 1921–1929 ω'_{un}: 1930–1939 $N = 9$	0.09 (0.09)	–	32.9(21) (0.05)

Notes: There are $I \times T = 350$ observations; y is real output growth, g is real government spending growth; y_u is growth in the US. The moment conditions apply for each decade and each country separately. N is the number of groups of countries. Estimation is by continuously updated GMM. The J_y-statistic tests the over-identifying restrictions.

Figure 1: Real Output per capita 1920-1939

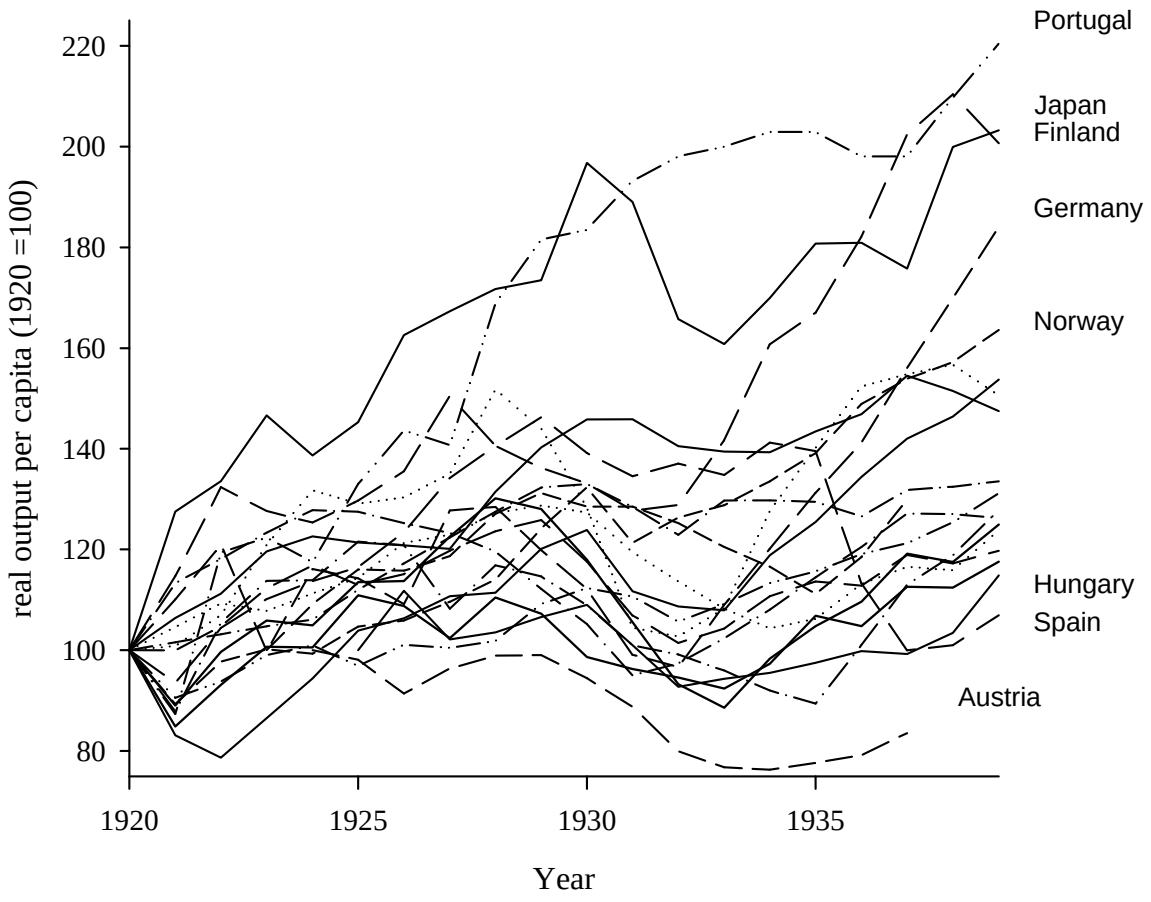


Figure 2: Real government spending per capita 1920-1939

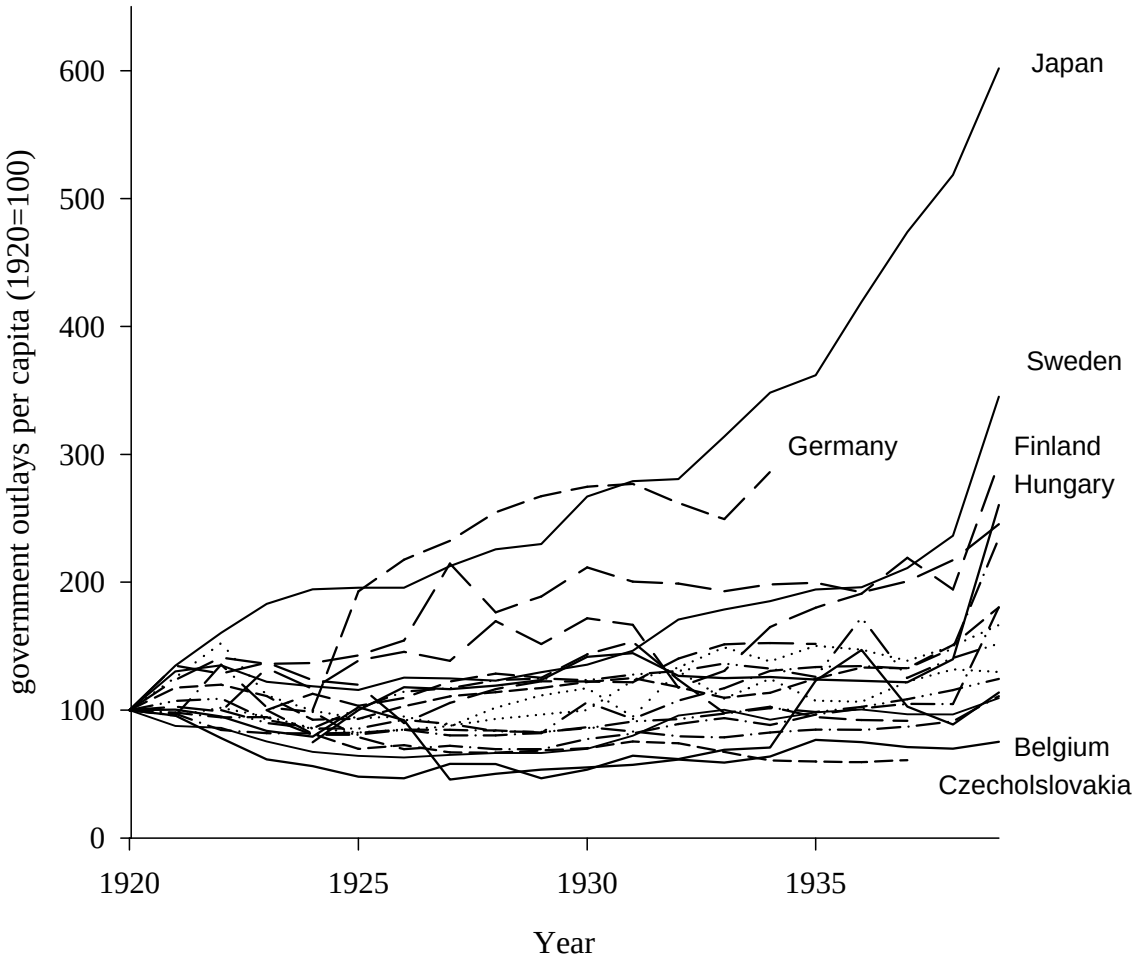


Figure 3: Government Spending Shares of Output

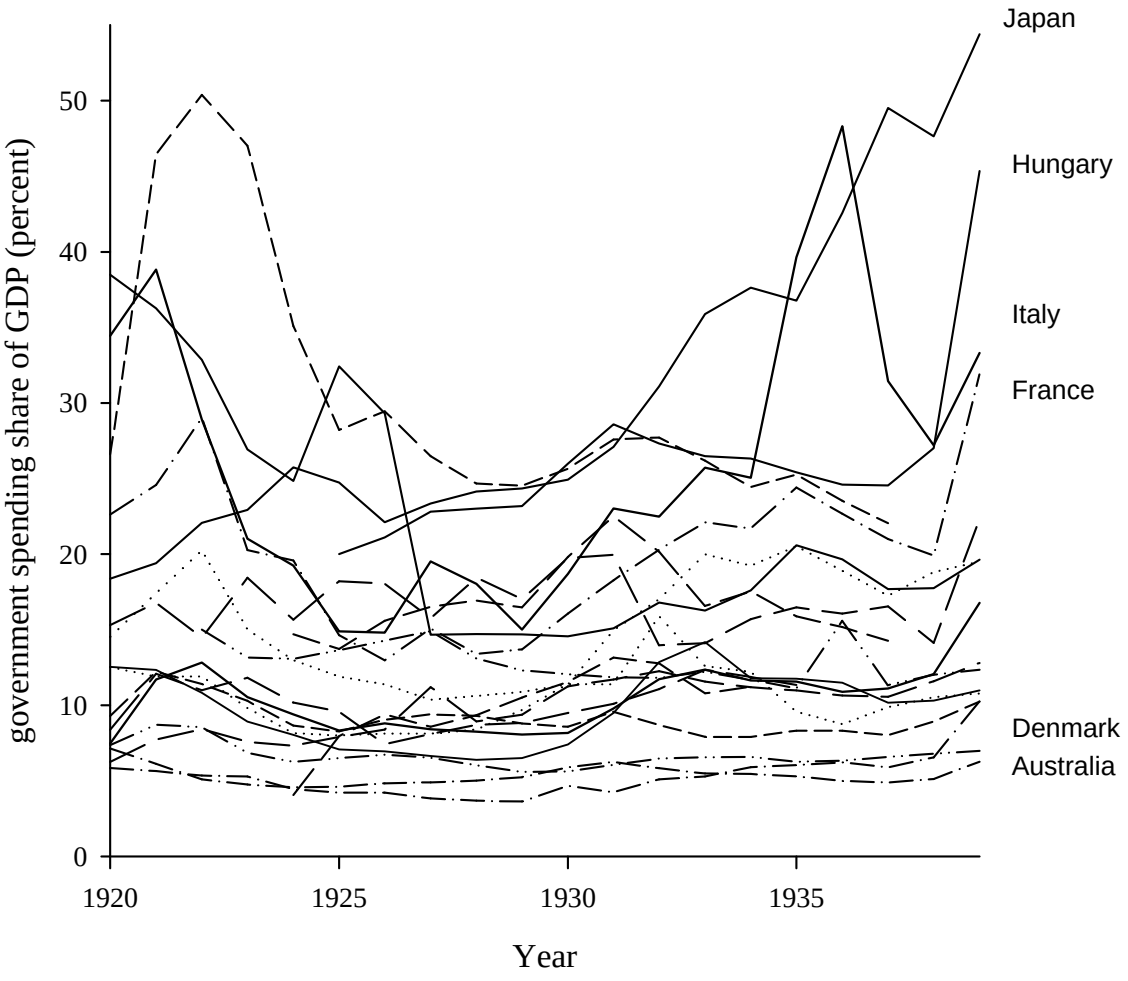


Figure 4: Fiscal policy responses and effectiveness

Countries $i = j$, k: identifying α_j and α_k, and β

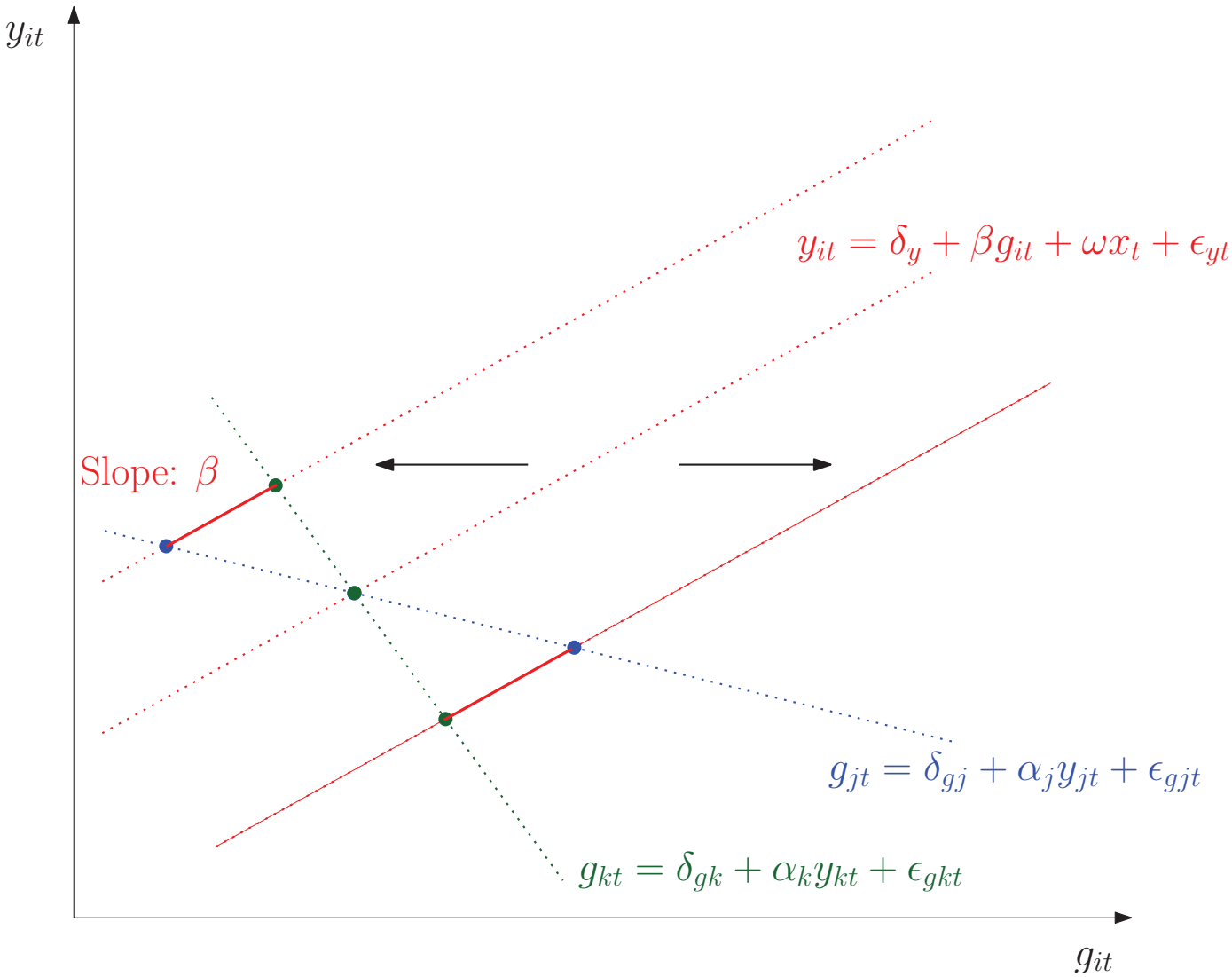
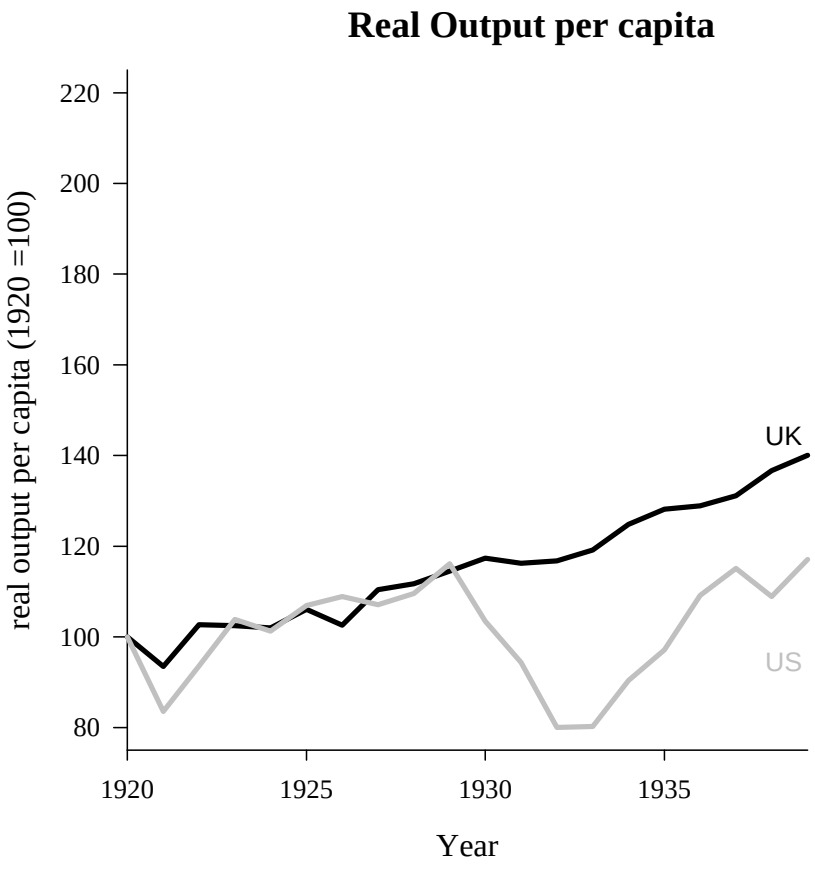


Figure 5: US and UK Data 1920-1939



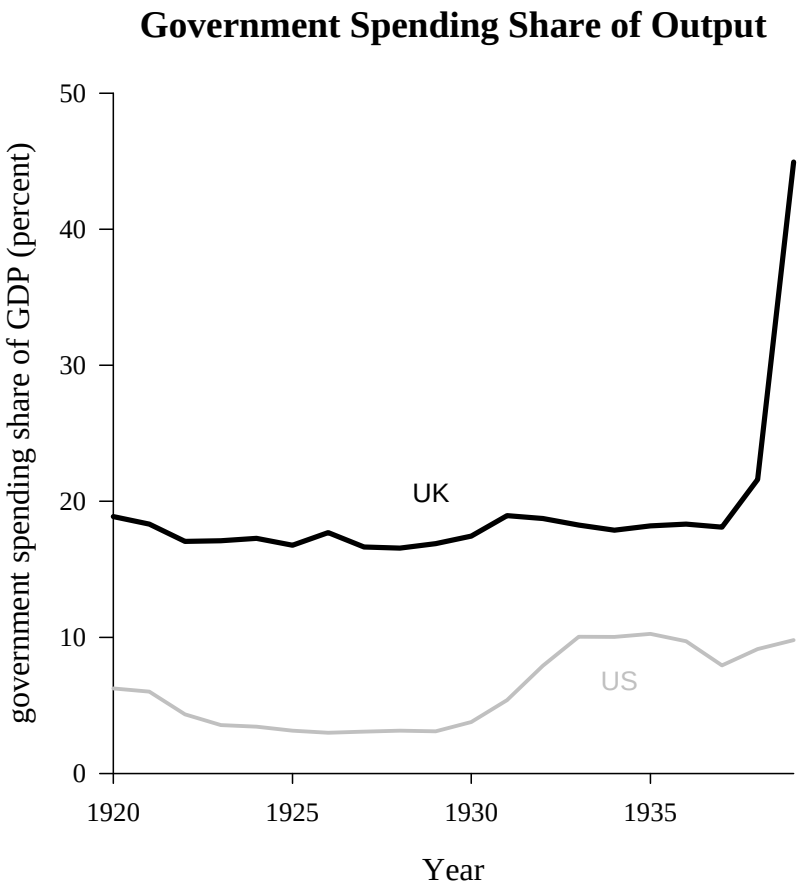


Figure 6: Gold Prices/Exchange Rates

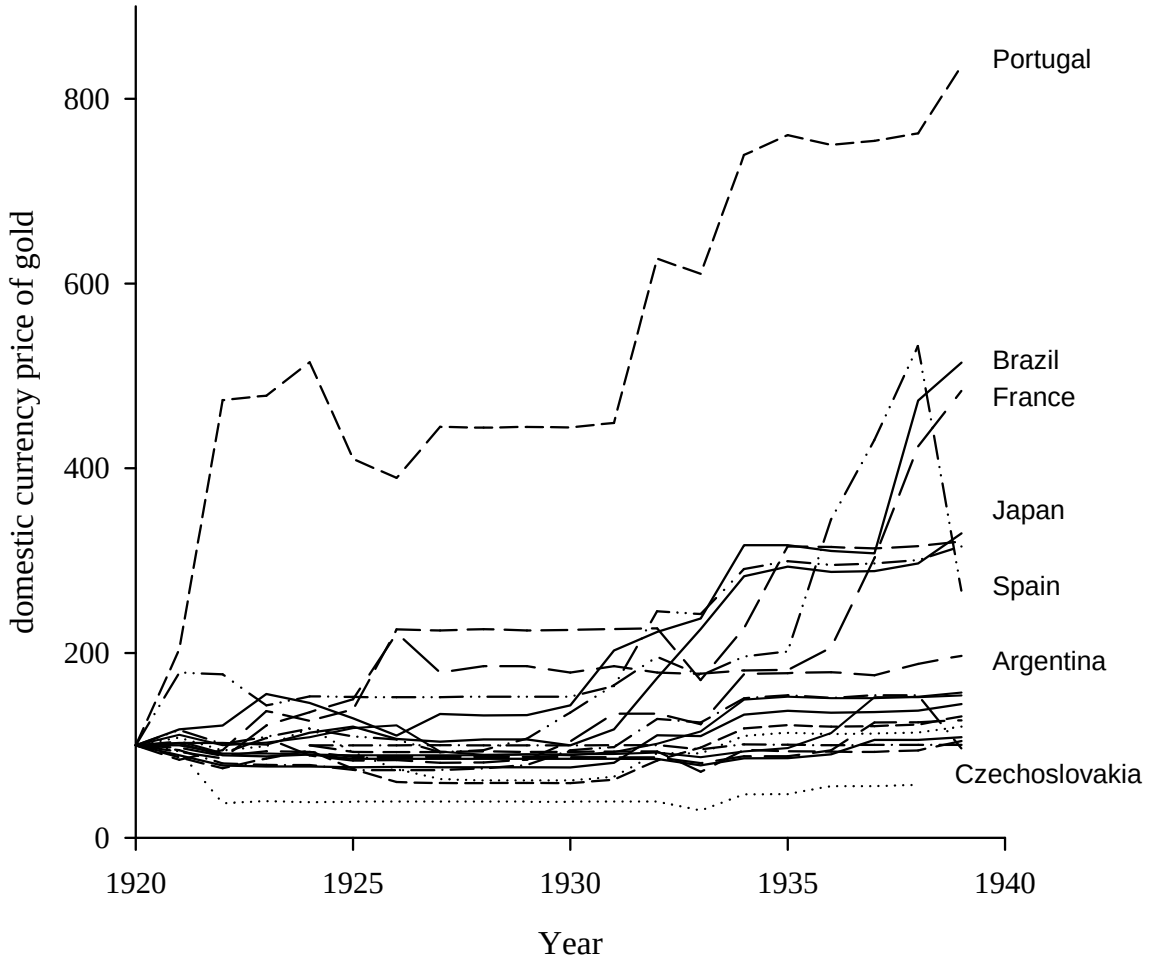
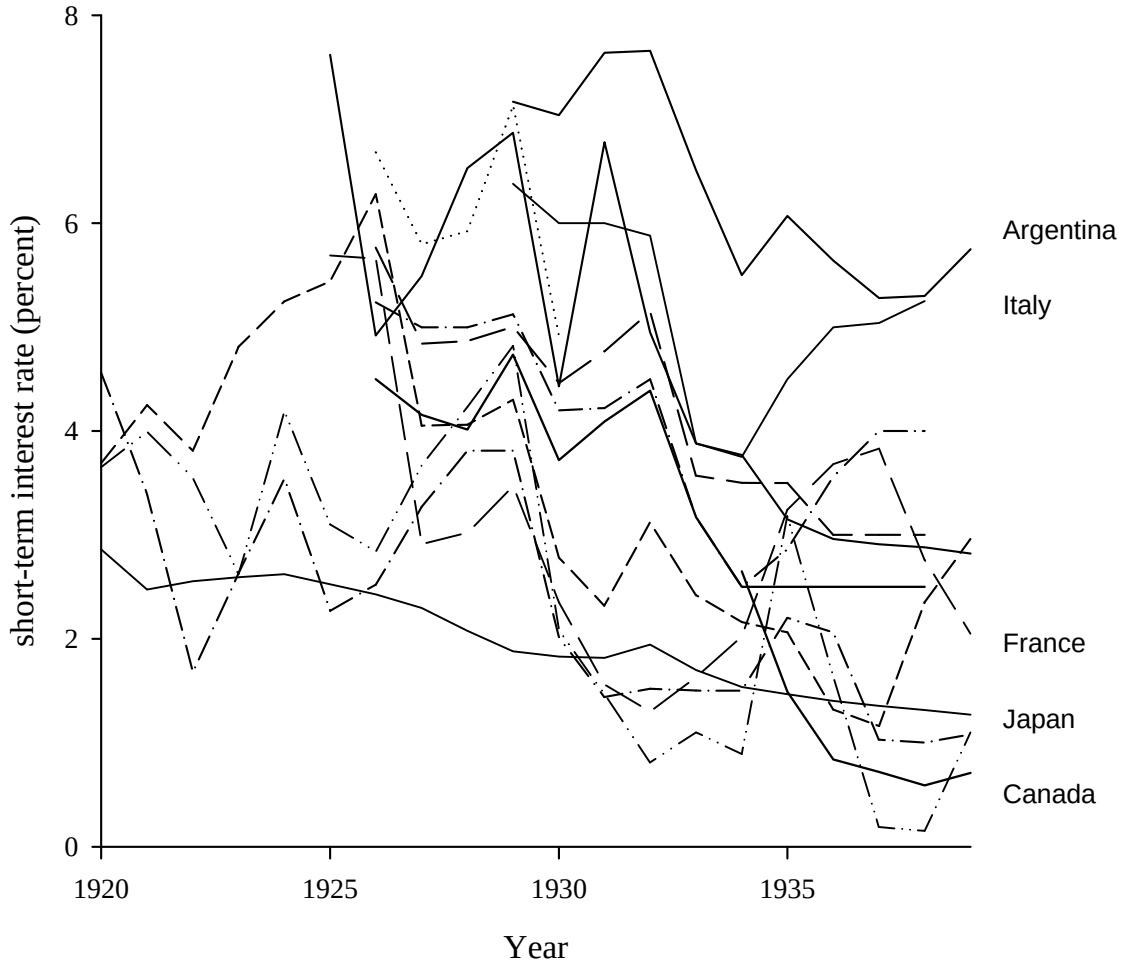



Figure 7: Short-Term Interest Rates