Is Deflation Costly After All? Evidence from Noisy Historical Data*

Daniel Kaufmann†

First version: 13 July 2016

This version: 28 November 2016

Abstract: I study the link between real activity and deflation, taking into account measurement problems in 19th century CPI data. Replications based on modern data show that measurement problems spuriously increase the volatility of inflation as well as the number of deflationary episodes, and they lower inflation persistence. As a consequence, estimates of the link between real activity and deflation may be attenuated because of the errors-in-variables problem. I find that real activity was on average substantially lower during 19th century deflations in the US, after controlling for measurement error using an IV-regression approach. Moreover, the average short-fall in real activity was not significantly different compared to the Great Depression. Using well-measured data for a panel of 17 industrialized economies shows that milder deflations were associated with a lower output gap. But, the association with GDP growth is not statistically significant.

JEL classification: E31, E32, N11, C36

Keywords: Deflation, real activity, measurement error, monetary history, IV.

---

* Most of this project was undertaken when I was visiting the Berkeley Economic History Laboratory (BEHL), whose hospitality I gratefully acknowledge. I thank an anonymous referee, Christiane Baumeister, Bernd Bartels, Gillian Brunet, Brad DeLong, Barry Eichengreen, Yuriy Gorodnichenko, Savina Gygli, Philipp Harms, Matthias Hölzlein, Florian Huber, Ronald Indergand, Dmitri Koustas, Tobias Renkin, Christina Romer, Gisela Rua, Samad Sarferaz, Jan-Egbert Sturm, Michael Siegenthaler, Eric Sims, Zach Stangebye, Richard Sutch, John Tang, Michael Weber, and Jonathan Wright, as well as seminar participants at UC Berkeley, the University of Notre Dame, the Federal Reserve Board, the University of Mainz, and the EEA-ESEM, for helpful comments and discussions. I also thank Samuel Williamson for granting permission to use the data from MeasuringWorth, as well as Steve Reed and Owen Shoemaker from the BLS for their valuable insights on historical and modern CPI data.

† KOF Swiss Economic Institute, ETH Zurich, Leonhardstrasse 21, CH-8092 Zurich, Switzerland. Phone: +41 44 632 04 45, e-mail: kaufmann@kof.ethz.ch

1. Introduction

Deflation is conventionally associated with low economic growth, high unemployment and a shaky financial sector. Keynes already argued in 1923 that moderate inflation is likely the lesser of two evils.[1] This view also resonates in the numerous nonconventional policy actions taken by modern central bankers since the Global Financial Crisis. Many of those actions are grounded on the fear that declining prices go hand in hand with dismal economic outcomes. However, there are theoretical and empirical reasons to believe that this fear is overblown. From a theoretical point of view, Friedman (1969) argued that optimal monetary policy can be characterized by a zero nominal interest rate and a moderately falling price level. Moreover, from an empirical point of view, the link between real activity and deflation appears to be weak (see e.g. Borio et al., 2015).[2]

Existing empirical studies often include pre-WWII data because the monetary regimes of the 19th and early 20th century brought about regular deflationary episodes (see Atkeson and Kehoe, 2004, Bordo and Filardo, 2005, Borio et al., 2015, and Eichengreen et al., 2016). Deflation was a necessary consequence of the metal-currency regimes that ensured long-term price-level stability instead of focusing on short-term stabilization policies (see e.g. Bernholz, 2003). During the 19th century US, the consumer price level declined nearly half of the time and the average annual deflation amounted to -4.7%. Therefore, deflation was not only frequent but also substantial.

This paper asks whether the lacking association between the 19th century deflations and real activity stems from measurement error in historical CPI data. Historical data often suffer from methodological deficiencies and measurement error (see Romer, 1986a, 1986b). Therefore, a relevant share of price-level changes may be artifacts of mismeasured macroeconomic data. If this is the case, the lacking association may stem from the well-known errors-in-variables problem (see Hausman, 2001). Intuitively, assume that deflation is actually associated with low GDP growth but we use a mismeasured CPI to classify deflationary and inflationary episodes. If we calculate the average GDP growth rate during deflations, some of those periods were in fact associated with rising prices and relatively high GDP growth. Therefore, the average growth rate based on the error-ridden classification will overestimate GDP growth rate during deflations. By contrast, if we calculate the average growth rate during inflations, some of them were actually associated with falling prices and low GDP growth.

---

[1] Keynes (1923), p. 40: "Thus Inflation is unjust and Deflation is inexpedient. Of the two perhaps Deflation is, if we rule out exaggerated inflations such as that of Germany, the worse; because it is worse, in an impoverished world, to provoke unemployment than to disappoint the *rentier*. But it is not necessary that we should weigh one evil against the other. It is easier to agree that both are evils to be shunned.".

[2] This tension between policy maker's views and the empirical evidence is nicely illustrated comparing a quote by Ben S. Bernanke (2002): "The sources of deflation are not a mystery. Deflation is in almost all cases a side effect of a collapse of aggregate demand—a drop in spending so severe that producers must cut prices on an ongoing basis in order to find buyers.", to a quote by Borio et al. (2015): "The evidence suggests that this link [between output growth and deflation] is weak and derives largely from the Great Depression.".

Therefore, we will underestimate average growth during inflations. Aigner (1973) shows that this classification bias—a variant of the attenuation bias due to classical measurement error—depends on the rate at which we misclassify deflationary as well as inflationary periods. Researchers acknowledge measurement error in historical price data as a caveat for their empirical results (see e.g. Barsky, 1987 and Benati, 2008). Little is known, however, how measurement error affects retrospective CPI estimates and to what extent it hampers measurement of the link between real activity and deflation.

This paper aims to fill this gap in three ways. First, it shows that measurement error is sizeable by replicating deficiencies of a popular 19th century US CPI based on modern post-WWII data. Second, a classic solution to the errors-in-variables problem is a proxy variable approach (see Hausman, 2001). I construct such a proxy based on wholesale price data and examine the link between real activity and deflation in an IV-regression framework for the US from 1800-1945. Third, I repeat the analysis with well-measured post-WWII data using a panel of industrialized economies.

The main findings may be summarized as follows. The replications of the measurement problems show more-frequent deflationary episodes, spuriously high volatility, and a lower persistence of inflation. In a worst-case-scenario, where all deficiencies apply at the same time, the standard deviation of inflation is twice as large compared to a correctly measured CPI. The higher volatility of inflation implies that many deflations are misclassified and, therefore, we may substantially overestimate real GDP growth during 19th century deflations. This is confirmed by the IV-estimates, which show that real activity was significantly and substantially lower during deflationary episodes. The IV-regressions suggest that an average deflationary episode was accompanied with about 4pp lower GDP growth and 9pp lower industrial production growth. Interestingly, there is no significant difference between deflationary episodes of the 19th century and the Great Depression. The panel regressions based on modern data provide additional, but weaker, evidence that deflation was associated with lower real activity. An association emerges only for the output gap, but not for GDP growth. This is in line with the fact that, compared to the 19th century US, the typical deflationary episode was less severe.

From a methodological point of view, this paper follows Romer (1986a, 1986b), Allen (1992), and Hanes (1998), who replicate methodological deficiencies in historical estimates of real activity, wages and wholesale prices using post-WWII data. The main contribution of the present paper is show that measurement issues are particularly relevant for historical CPI data and to gauge the implied attenuation and classification biases. From a substantive point of view, the paper is closely related to Atkeson and Kehoe (2004), Bordo and Filardo (2005), and Borio et al. (2015), who find only a weak link between real activity and deflation for sizeable panels of countries and, in particular, when excluding the Great Depression. Eichengreen et al. (2016), however, report that the link becomes more pronounced when they use wholesale prices instead of consumer prices. The explanation for those conflicting results of this paper focuses on mismeasured CPI data. The errors-in-variables problem also attenuates measures of inflation persistence. Therefore, the paper is related to a large body of literature finding little inflation persistence during the 19th century (see Klein 1975, Shiller and Siegel, 1977, Sargent 1973, Barsky, 1987, Barsky and DeLong, 1991, and Benati, 2008) and examining the cyclicality and flexibility of prices and wages during the 19th and early 20th century using Phillips-curve-type specifications (see Cagan, 1975, Sachs, 1980, Gordon, 1980, and Hanes, 1998).

In what follows, I first demonstrate the impact of measurement error in two different regression frameworks. Then, I propose three ways to recover the actual association between real activity and deflation and present the empirical results. After discussing various robustness and specification tests, the last section concludes.

2. The errors-in-variables problem

The errors-in-invariables problem hampers estimating the state of the real economy during deflationary episodes. I first discuss the problem in a widely-used reduced-form regression framework. Researchers have examined the link between real activity and deflation regressing GDP growth on a deflation indicator (see Borio et al., 2015 and Eichengreen et al., 2016).[3] In the simplest case of only one country and no additional control variables, the regression equation reads:

$$y_t = c + \delta 1\{\pi_t < 0\} + \varepsilon_t, \tag{1}$$

where y_t is a measure of real activity and $1\{\pi_t < 0\}$ is a dummy variable that equals unity if inflation is negative and zero otherwise. The error term ε_t captures unexplained factors including independent measurement error in the real activity variable. A negative coefficient on the deflation dummy indicates that real activity has been on average lower during deflationary episodes than during inflationary episodes.

If the analysis is based on a mismeasured inflation rate, for example $\tilde{\pi}_t = \pi_t + \omega_t$, the resulting dummy $1\{\tilde{\pi}_t < 0\}$ classifies some periods as deflations, when prices were actually rising, and some periods as inflations when prices were in fact falling. We can decompose the correctly measured but unobserved deflation indicator into the error-ridden indicator, an indicator for misclassified deflation periods, and an indicator for misclassified inflation periods:

$$1\{\pi_t < 0\} = 1\{\tilde{\pi}_t < 0\} - 1\{\tilde{\pi}_t < 0, \pi_t > 0\} + 1\{\tilde{\pi}_t > 0, \pi_t < 0\}. \tag{2}$$

If we insert the decomposition into equation (1), the regression equation reads:

[3] Measurement error in historical price data biases inflation persistence and slope coefficients in equations using CPI inflation as a right-hand-side variable. Therefore, measurement error likely affect structural VARs estimated on historical data, for example, along the lines of Bayoumi and Eichengreen (1996) and Bordo and Redish (2004). Going beyond reduced-form regressions and examining the impact on structural VAR analysis is beyond the scope of this paper.

$$y_t = c + \delta 1\{\tilde{\pi}_t < 0\} + v_t, \quad v_t \equiv -\delta 1\{\tilde{\pi}_t < 0, \pi_t > 0\} + \delta 1\{\tilde{\pi}_t > 0, \pi_t < 0\} + \varepsilon_t. \qquad (3)$$

The OLS estimate of δ in equation (3) suffers from a classification bias because the regressor $1\{\tilde{\pi}_t < 0\}$ will be negatively correlated with the error term v_t through the unobserved misclassified deflationary and inflationary episodes. Aigner (1973) shows that the OLS estimate converges in probability to

$$plim\ \hat{\delta}_{OLS} = \delta(1 - \eta^- - \eta^+), \qquad (4)$$

where η^- and η^+ denote the share of misclassified deflationary and inflationary periods, respectively. The misclassification factor $(1 - \eta^- - \eta^+)$ equals unity if we classify both, deflationary periods as well as inflationary periods, correctly.

The bias has an intuitive interpretation. Assume that deflation is actually associated with low GDP growth but we use a mismeasured CPI to classify deflationary and inflationary episodes. If we calculate the average GDP growth rate during deflations, some of those periods were in fact associated with rising prices and relatively high GDP growth. Therefore, the average growth rate based on the error-ridden classification will overestimate GDP growth rate during deflations. By contrast, if we calculate the average growth rate during inflations, some of them were actually associated with falling prices and low GDP growth. Therefore, we will underestimate average growth during inflations.

The classification bias is a variant of the well-known attenuation bias (see Hausman, 2001). When a continuous right-hand-side variable is measured with classical measurement error, the OLS estimate will be biased towards zero. If we regress a measure of real activity on inflation, the OLS estimator of the slope coefficient converges to the true coefficient multiplied by the relative variances of the actual and mismeasured inflation rates:

$$plim\ \hat{\beta}_{OLS} = \beta \sigma_{\pi}^2 / \sigma_{\tilde{\pi}}^2. \qquad (5)$$

The attenuation factor declines, and therefore the bias increases, if the variance of the error-ridden inflation rate rises relative to the variance of the actual inflation rate. The same attenuation factor as in equation (5) carries over to the OLS estimate of the slope coefficient in a first-order autoregressive model (see Staudenmayer and Buonaccorsi, 2005) and, therefore, measurement error attenuates measures of inflation persistence. Intuitively, regressing inflation on its lag is a special case of a regression with measurement error in a continuous right-hand-side variable.

3. Revisiting deflation and depression

Against the backdrop of the reduced-form regression framework, this paper addresses the measurement error problem in three ways. First, it quantifies the measurement error variance as well as the share of misclassified episodes by replicating deficiencies of retrospective CPI estimates of the 19th century US. Because we observe both, the correctly measured and the error-ridden CPI, we can calculate the misclassification factor as well as the attenuation factor. This allows to gauge the size of the bias and therefore, whether the errors-in-variables problem is relevant for historical studies on the link between real activity and deflation. Second, to estimate the size of the association, not only the potential bias, we have to estimate equation (3) using historical data and control for measurement error.[4] To resolve the bias, I use an IV-regression approach (see Hausman, 2001). For this strategy to work, the instrument has to be correlated with the error-ridden CPI but uncorrelated with the measurement error. I therefore calculate a proxy for US CPI inflation from 1800-1890 based on wholesale price data.[5] This proxy is then used to instrument the error-ridden indicator in IV-regressions of equation (3). As a third solution, we can repeat the analysis with modern data that are measured more accurately. To this end I use the historical data set by Jordà et al. (2016) for the post-WWII era. The data set comprises 17 countries and regularly used control variables. The disadvantage of this data set is that deflations were less frequent and more benign than during the 19th century.

3.1. Assessing the size of bias

To assess the potential size of the bias, I replicate the methodological deficiencies of a popular composite CPI during the 19th century (Officer and Williamson, 2016a).[6] The composite index covers the period 1774-2015 and combines a careful selection of alternative retrospective estimates. Although the entire series most likely represents the best available estimate at any given point in time, the various segments suffer from important methodological deficiencies. They can be traced back to scarce retail price data, especially during the 18th and 19th century. It is thus not surprising that the retrospectively estimated segments are more strongly affected by measurement error than the post-WWII segments for which actual retail price surveys have been conducted.

[Table 1 about here]

Table 1 lists examples of the most important methodological deficiencies in the composite CPI. First, David and Solar (1977) use wholesale prices to approximate prices at the retail stage. Second, price indices for small geographical areas are often used to represent prices for the US as a whole. For

---

[4] Even though the bias may be substantial, the association may be economically irrelevant if the true $\delta = 0$.

[5] Historical wholesale prices have the additional advantage that they are regarded to be more accurately measured than consumer prices. Therefore, the signal to noise ratio should be higher and the attenuation bias smaller. But even if the measurement error in wholesale prices would be as large as measurement error in consumer prices, wholesale prices are more volatile than consumer prices. A deflation signal from wholesale prices may therefore be more accurate than a deflation signal from consumer prices and classification bias may be again smaller. This may be one explanation why Eichengreen et al (2016) found a significant link using wholesale prices in contrast to Bordo et al. (2015) using consumer prices.

[6] This choice does not imply that this index is particularly subject to measurement error. On the contrary, I chose this index because it reflects a careful selection of various segments and the properties of the index are well documented by Officer (2014). This paper draws repeatedly on his careful description of alternative retrospective estimates of US CPI inflation.

the period before 1851, researchers regularly use the indices constructed by Adams (1939), which are based on retail prices paid by farmers from Vermont. A third deficiency is the small number of individual price quotes that are used to construct the price indices. In one of the most comprehensive surveys on retail prices for the 19th century—the so-called Weeks Report—the number of observations is much smaller than in modern surveys.[7] Retail price data become even scarcer from 1880 to 1890, after the Weeks Report ended and before the U.S. Bureau of Labor Statistics (BLS) started to collect retail prices for food items on a broader scale (see Officer, 2014). Fourth, some price indices are available only for specific periods and have to be interpolated in between. Long (1960) approximates the prices of several items, including rent, by a linear interpolation over the entire 1880s. Fifth, information on rent for housing is scarce. Lebergott (1964) constructs a reproduction cost index by equally weighting the cost of construction materials and wages for low-skilled workers. Sixth, a general defect is the lack or minimal coverage of service prices. For example, the indices by Lebergott (1964) and Hoover (1960) comprise only few service items: rent, shoe repairs and physician fees paid by Vermont farmers.[8]

Because retail price data are particularly scarce for the 19th century, the methodological deficiencies likely lead to more-serious measurement error in a CPI than in a wholesale price index. To quantify the impact of those measurement errors on the time-series properties of CPI inflation, I replicate the deficiencies using modern CPI data. The replications are based on several special aggregates underlying the BLS CPI and begin in 1956, when the BLS started reporting service prices on a monthly basis.[9] For simplicity, the replications are constructed as Laspeyres-type indices with constant expenditure weights at an annual frequency. The weights for 1869 (representing the 19th century CPI) and 2013 (representing the post-WWII CPI) stem from Gordon (2015). During the 19th century, the consumption basket was heavily tilted towards nondurables, particularly food items. Almost 70% of the budget was spent on nondurable goods, whereas only 20% of the budget was spent on services.[10] By 2013, the expenditure shares for the two commodity groups reversed, whereas the expenditure share on durable goods has not changed significantly. Although this change in the consumption basket does not represent a deficiency as such, it also affects the time-series properties of CPI inflation by attaching more weight to deficiencies particular to non-durable goods prices.

Modern BLS data allow direct replication of five of the six deficiencies listed in Table 1. An index for Philadelphia replicates limited geographical coverage.[11] Then, I construct a CPI based on the

---

[7] In *1880 Census of the United States, Vol. xx, Joseph D. Weeks, Report on the Statistics of Wages in Manufacturing Industries, with Supplementary Reports*.
[8] Similar issues plague CPI estimates well into the 20th century. It was not until 1940 that the CPI began to be published on a monthly basis (although many service prices were still collected only quarterly). Before, the CPI was only published for irregular intervals or even only for December. From 1913-1921, the BLS retrospectively estimated a monthly CPI, interpolating prices for many items that were not collected monthly (see Officer, 2014).
[9] The sources for all series are given in Appendix A.
[10] The table is reproduced in Appendix A.
[11] Results for other regions (Northwest, Midwest, South, West) are similar but not reported because the series are available only for a shorter sample.

special aggregates nondurables, durables and services, replacing the indices for nondurables and durables by their counterparts in the producer price index. Note that this represents a worst-case scenario, in which consumer prices are directly replaced without adjusting for the different volatility of the two series. For replicating linear interpolation, I keep only every 10th annual observation of the index for shelter (in addition to the first and last observation of the series), linearly interpolate the missing values in between, and calculate the aggregate with the all-item less shelter index. The reproduction cost index replaces the shelter index by an equally weighted average of producer prices for construction materials and wages for low-skilled workers from Officer and Williamson (2016b). Then, the CPI less services replicates the lack of service prices.

The sixth deficiency is the small number of individual price quotes used to construct retrospective CPI estimates. In a typical price index for the 19th century the number of price quotes ranges from about 2,000 to just over 8,000. This range represents two scenarios based on the discussion by Hoover (1960) on the number of missing values in the Weeks Report (see Appendix B for details). Nowadays, the number of individual price quotes underlying annual CPI inflation rate amounts to more than 1,000,000. Because the BLS reports the sampling standard error for the modern CPI inflation rate, we can gauge the sampling error for a smaller underlying sample. This requires two simplifying assumptions. First, assume that the CPI inflation rate is the unweighted average of individual price changes and second, that those individual price are i.i.d. with finite variance s^2. Then, a central limit theorem applies according to which the unweighted average converges to a normal distribution, with expected value equal to the true inflation rate and a sampling variance $\sigma^2 = s^2/N$. Because the BLS publishes an estimate of the sampling standard error (σ) as well as the number of observations (N), this allows us to gauge the sensitivity of the sampling error to a reduction of the number of observations, holding constant the variance of individual price changes (s^2). The annual sampling standard error for the CPI inflation rate in 2014 amounts to 0.07 percent (see Shoemaker, 2014 and Appendix B). In the replication, I assume that the number of observations lies at the upper range for a historical price index (8,000 observations). This increases the sampling standard error to 0.78 percent and introduces substantial uncertainty into CPI inflation measurement: A 95% confidence interval for the measured CPI inflation rate of 1% would amount to [-0.5, 2.5].

[Table 2 about here]

Before discussing the replications, Panel (A) in Table 2 shows descriptive statistics of the composite CPI inflation rate over various subsamples starting in 1800. The sample is split before WWI and after 1956 for comparability with the replications.[12] The standard deviation of inflation was more than twice as large before WWI as after WWII. Moreover, we observe a strong increase in the

[12] A sample split according to different monetary regimes during the 19th century as defined by Bordo and Kydland (1996) and for other inflation series is given in Appendix A.

persistence of inflation in the post-WWII era. Finally, the price level declined almost half of the time before WWI, whereas deflation became an anomaly after WWII. Those stylized facts are in line with the findings of Barsky (1987) and Bordo and Filardo (2005). In what follows, the replications show how those changes in the time-series properties of inflation may be related to measurement issues in historical CPI data.

The first two replications show the impact of assuming constant weights and applying 19th century expenditure weights for comparability.[13] Replications (3)-(9) each mimic a particular methodological deficiency. Because of limited geographical coverage, the persistence falls to 0.80. As a result of using nondurable wholesale prices, the volatility of inflation increases and the persistence declines as well. There is no difference, however, if we approximate durable goods prices by their wholesale stage counterparts. Replications (6)-(7) show the impact of linear interpolation of rent and using a reproduction cost index. Those deficiencies actually lead to a lower volatility. Also, the impact on the persistence is small for the linear interpolation. The lack of service prices leads to a more pronounced change. We observe both a decline in persistence and an increase in volatility. Finally, adding sampling error introduces classical measurement error.[14] We know from econometric theory that the well-known attenuation bias of the errors-in-variables problem carries over to autoregressive models (see Staudenmayer and Buonaccorsi, 2005). This is indeed what we observe and the persistence of inflation declines to 0.79.

The remaining columns of the table indicate to what extent measurement error leads to spurious deflations and reports the implied attenuation and classification factors. The largest increase in spurious deflations we observe for missing services. But also, most other deficiencies increase the share of deflations slightly. Because of the higher variance, the attenuation factor falls below unity when using the PPI for non-durable goods, because services prices are missing, and because of classical measurement error due to the small sample. But, the attenuation bias is relatively small. The classification bias, however, is more relevant. Because we observe only a small share of deflations in the actual data, and most deficiencies increase the share of periods with falling prices, the share of misclassified deflations increases as well. For the case of missing services prices, for example, the misclassification factor implies that the OLS estimate of equation (3) would amount only to one third of the true coefficient.

---

[13] Replication (1) in Panel (B) is based on the special aggregates nondurables, durables, and services, using weights from 2013. We see that the simplifying assumption of constant weights does not materially affect the descriptive statistics. Applying expenditure weights from 1869 to durables, nondurables and service prices has a stronger impact. The persistence falls to 0.70, and the standard deviation increases to 2.8 percent. This stems from the fact that inflation for non-durable goods is particularly volatile and less persistent than service price inflation.

[14] The descriptive statistics are means of 10,000 simulations, adding independent normally distributed measurement error with sampling error variance adjusted by the relative number of observations in modern and historical price data.

Individually, the methodological deficiencies have modest effects on the time-series properties of CPI inflation. A combination of the deficiencies, however, lead to relevant differences.[15] Panel (C) presents worst-case-scenarios where all deficiencies apply at the same time: they combine the impact of wholesale prices, linear interpolation, reproduction cost index, lack of service prices, and sampling error at weights for 2013 and 1869, respectively.[16] The standard deviation roughly doubles compared to the actual CPI inflation rate. This is also reflected in the attenuation factor which falls to 0.2. As a consequence, also the persistence drops substantially to 0.28 and 0.42, at 2013 and 1869 weights, respectively. Again, the classification bias is more severe than the attenuation bias. The misclassification factor implies that the OLS estimate may be up to 10 times smaller than the true coefficient. Interestingly, this result holds largely independent of the composition of the consumption basket.

How do those simulations compare with the actual time-series properties of the composite CPI inflation rate in Panel (A)? The standard deviation of CPI inflation was 5.7 percent before 1914 and 2.6 percent after 1956. This difference is only slightly larger than what the combined deficiencies imply. The measured persistence rises from 0.43 before WWI to 0.85 after 1956. This is perfectly in line with combined deficiencies at weights from 1869. The attenuation factor of 0.2 would even imply a decline in persistence from 0.85 to 0.2. Note, however, that the attenuation factor only applies in the case of classical measurement error. Finally, the share of deflationary episodes falls from 0.46 before WWI to almost zero after 1956. This is driven to a substantial extent by the higher average inflation rate and therefore this difference can be unlikely traced back to the deficiencies investigated in this paper. Nevertheless, the high volatility of inflation because of the methodological deficiencies implies that many deflations during the 19th century may be artifacts of mismeasured CPI data.

To summarize, this section has shown that methodological deficiencies and a small number of individual price quote observations increase the volatility and reduce the persistence of inflation. Moreover, measurement error artificially increases the number of deflationary episodes. Because we falsely classify inflationary periods as deflationary episodes, studies examining the link between GDP growth and deflation will therefore suffer from a classification bias which attenuates the actual link between real activity and deflation.

3.2. Resolving the bias using IV

The errors-in-variables problem can be resolved using an independent proxy variable to instrument the error-ridden CPI (see Hausman, 2001). For the 19th century, such a proxy can be

[15] An augmented-Dickey-Fuller test does not reject the null of a unit root for the actual CPI inflation rate from 1957-2015. However, for the combined replications the test rejects the unit root hypothesis at the 1% level. Moreover, the median-unbiased 90% confidence interval for the persistence by Hansen (1999) amounts to [0.15, 0.61] and therefore clearly excludes unity.

[16] Limited geographical coverage could not be replicated because not all subindices are available for Philadelphia on a sufficiently long sample.

constructed based on wholesale prices from Warren and Pearson (1933). Although this proxy shares some of the methodological deficiencies with the composite CPI, it stems from a different data source and should therefore be a valid instrument to control for classical measurement error.

The CPI proxy is based on wholesale prices for the commodity groups “food”, “textile products”, “fuel and lighting”, and “house furnishings”, which are aggregated to a Laspeyres-type index using expenditure weights by Gordon (2015).[17] Those commodity groups cover approximately 70% of the expenditure weights in 1869. The most important missing item, making up 18% of the consumption basket, is rent. Moreover, because house furnishings prices are not available before 1840 I interpolate the series with the weighted average of the other available price series.

[Figure 1 about here]

Panel (A) of Figure 1 shows the composite CPI inflation rate as well as the proxy based on wholesale prices from 1800-1890. The two series display reasonably similar turning points and inflationary and disinflationary episodes. Because the proxy is constructed using wholesale prices, it is more volatile. The correlation between the two series, however, is substantial, suggesting that the proxy is a relevant instrument. Interestingly, despite their high correlation, the two variables give different signals concerning deflationary or inflationary episodes. In 26% of all years, the two indicators do not agree on whether it was a deflationary or inflationary episode. This share is relatively stable for various subsamples.

For the CPI proxy to be a valid instrument, its measurement error has to be uncorrelated with the measurement error in the composite CPI. As a necessary condition, it therefore has to be based on different data sources than the individual segments of the composite CPI by Officer and Williamson (2016a). The Warren and Pearson (1933) data stem from New York newspapers supplemented by prices published in the U.S. Finance Report for 1863 (see Hanes, 2006).[18] By contrast, from 1800 to 1851, the composite CPI uses retail prices for some benchmark years and prices paid by Vermont farmers to interpolate in between. From 1851 to 1860, it is partly based on wholesale prices for fruits. However, the sources are distinct: Hoover (1960) uses prices for Philadelphia and from the so-called Aldrich Report.[19] The Lebergott (1964) segment is again based mainly on the Weeks Report, and the only wholesale prices used are for building materials in the reproduction cost index (which are not used to construct the proxy). From 1880 to 1890, the segment by Long (1960) is based on thin and sketchy retail

---

[17] See Appendix A for data sources. The Warren and Pearson (1933) commodity groups are matched with the weights from Gordon (2015) as follows: “foods” with “food, alcohol for off-premises consumption”; “textile products” with “clothing and footwear” as well as “dry goods for making clothing at home”; “fuel and lighting” with “tobacco, printed material, heating/lighting fuel”; and “house furnishing goods” with “furniture, floor coverings, house furnishings”.

[18] *Report of the Secretary of the Treasury on the State of the Finances* (38th Congress, 1st Session, 1863).

[19] *Wholesale Prices, Wages, and Transportation* (Senate Committee on Finance, 52nd Congress., 2nd Session, Report 1394, Part 2, 1893).

data because it refers to the difficult period after the Weeks Report. There is no indication that wholesale prices were used. After 1890, the segment by Rees (1961) uses wholesale prices for eleven items from the BLS (1923). The Warren and Pearson (1933) data end in 1890, and the longer series provided by Hanes (1998) are based on the same BLS data. I therefore calculate the proxy only for the period until 1890, for which, to the best of my knowledge, the wholesale price data used to construct the CPI proxy do not stem from the same source as the data underlying the composite CPI.

As another identifying assumption, we require that the noisy proxy shares the actual inflation rate as a common trend with the error-ridden CPI. Although we cannot test this assumption, it is possible to construct the proxy using modern PPI data and compare it to the well-measured modern CPI inflation rate.[20] This proxy covers only 13.7% of the consumption basket in 2013. Nevertheless, Panel (B) of Figure 1 shows that it is correlated with CPI inflation and reflects major up- and downturns. Moreover, a regression of the CPI inflation rate on the proxy yields a coefficient of 0.4, which is statistically significant at conventional significance levels with an R-squared of 0.5. This suggests that it is reasonable to assume that the proxy is also informative about inflation in the historical data, when the goods included in the proxy covered a larger share of the consumption basket.

In what follows, I estimate variants of equation (3) using a deflation dummy based on the CPI inflation rate from Officer and Williamson (2016a). In the IV-regressions, the first stage instruments the deflation dummy by a corresponding dummy based on the proxy. This procedure includes nonlinear terms of the instrument in the second-stage regression.[21] Specification tests of the first-stage regressions for all IV-regressions are given in Appendix C. For the baseline case with no additional control variables, we may consult the rk *LM*-statistic tests whether the model is underidentified in the presence of heteroscedasticity and autocorrelation (Kleibergen and Paap, 2006). The null is rejected at common significance levels. The rk *F*-statistic tests whether the model is only weakly identified. The statistic amounts to 31.9. The 5% critical value, for testing whether the asymptotic bias due to a weak instrument exceeds 10% of a worst-case benchmark amounts to 23.1 (Montiel Olea and Pflueger, 2013, derive critical values for the case of HAC-robust standard errors). This suggests that the dummy instrument is strong and confirms the visual impression from the continuous variable in Figure 1. This is the case for most IV-regressions reported in this paper. Therefore, these statistics are in the following only discussed for specifications we may worry that the instrument is weak.

[Table 3 about here]

---

[20] See Appendix A for data sources. The BLS PPI commodity groups are matched with the weights from Gordon (2015) as follows: “Processed foods and feeds” with “food, alcohol for off-premises consumption”; “Apparel” with “clothing and footwear”; “Fuels and related products and power” with “tobacco, printed material, heating/lighting fuel”; and “Textile house furnishings” with “furniture, floor coverings, house furnishings”.
[21] Alternatively, I followed Wooldridge (2002), p. 237, and regressed the CPI inflation rate on the instrument and control variables and then obtained the fitted values. Afterwards, I included a dummy based on the fitted values as an instrument. The results remained similar and are therefore not reported.

Table 3 shows the results for various measures of real activity:[22] Real per capita GDP growth, industrial production growth, both in percent, and percentage deviations of the two variables from their trends.[23] The estimates are alternatively based on OLS and IV for the time period 1800-1890. Panel (A) shows that OLS estimates for GDP and industrial production growth are not statistically different from zero. This supports the view that the link between deflation and real activity is weak when excluding the Great Depression. Using the proxy deflation dummy as an instrument, however, the estimated coefficients increase in size and become statistically significant at least at the 10% level. A deflationary episode coincided on average with 3.6pp lower GDP growth and 8.2pp lower industrial production growth. For the gap measures, we also observe that the IV-regressions yield a more strongly negative and statistically significant association. The ratio between the IV and OLS coefficients gives us an idea of the implied misclassification factor. The IV estimates are larger by a factor of 2 to 3, depending on the specification and real variable in question. The corresponding misclassification factor therefore amounts to 0.3 to 0.5, which is somewhat larger than what the combined replications imply, but in line with the individual replications. Qualitatively, the result remains unchanged in Panel (B) when controlling for equity price changes as well as major banking crises (Jalil, 2015).

Existing empirical studies stress that the Great Depression was an exceptional episode and that most other deflations were more benign. To assess this finding against the backdrop of mismeasured CPI data, we need an instrument covering the longer sample including the Great Depression. I use a composite WPI inflation rate based on data from Warren and Pearson (1933), Hanes (1998) and BLS after 1913. From 1800-1890, the WPI inflation rate is highly correlated with the proxy and the R-squared in a linear regression amounts to 0.94. Note that it is difficult to strictly establish that the data sources between the WPI and the composite CPI do not overlap because Rees (1961) occasionally used wholesale prices from 1890-1914. However, all results are robust to excluding the Rees (1961) segment from the analysis.

[Table 4 about here]

The results are shown in Table 4. On the entire sample from 1800-1945, the OLS coefficients are statistically significant, supporting the view that the period including the Great Depression is largely responsible for the significant association. Still, the IV-regressions yield substantially larger coefficients than OLS. The implied misclassification factors, dividing the OLS-coefficients by the IV-coefficients,

---

[22] Classical measurement error in the left-hand-side variable only reduces the precision but does not bias the OLS estimator. I still examine various measures of real activity to account for the possibility of non-classical measurement error.

[23] Real per capita GDP stems from Johnston and Williamson (2016) and industrial production from Davis (2004). The GDP series is already linked with modern data sources. Davis' series ends in 1914, and the official industrial production series starts only in 1919. I bridge this gap using the manufacturing production series by Fabricant (1940). Following Davis et al. (2009), the trends are estimated using a Hodrick-Prescott-filter with the smoothing parameter set to 100.

range from 0.2 to 0.5. To test whether the Great Depression was indeed significantly different, Panel (B) includes an interaction term with a dummy for the post-1914 period. For GDP growth, the OLS coefficient on this interaction term is significant and sizable. This results changes using IV. Deflation is associated with 5.7pp lower GDP growth over the entire sample and the interaction term including the Great Depression period does not significantly differ. For all IV-regressions, the coefficient for the entire period increases in size and turns statistically significant and, by contrast, the coefficient on the period including the Great Depression turns insignificant. This suggests that the deflations during the 19th century were, in terms of real activity, similar to the Great Depression after accounting for measurement error.

[Table 5 about here]

So far, we have not taken into account the severity of deflations. A deflationary episode with a minor drop in the price level was treated equally to a severe deflation with substantially falling prices. Table 5 shows an alternative specification, where real activity is regressed on inflation, the deflation dummy, and an interaction term. The regressions show whether real activity is significantly associated with CPI inflation and whether the association is stronger when CPI inflation is negative. Therefore, a positive coefficient on the interaction term implies that disinflation, when prices are already falling, is associated with a stronger decline in real activity. I instrument this interaction term with the corresponding interaction term based on the proxy and wholesale prices, respectively.

Panel (A) shows the results using the instrument based on the proxy variable. The results from this specification should be discounted, however, because the rk *F*-statistic is lower than the 5% critical value suggesting that the IV-estimates may be substantially biased. Moreover, using IV, the interaction term is imprecisely estimated and hardly statistically significant. Extending the analysis to a larger sample using instruments based on the WPI yields more reliable results.[24] Panel (B) shows that the IV-estimate is always statistically significant and larger than the OLS estimate. Finally, Panel (C) tests whether the interaction term is significantly different during the Great Depression. Using OLS, we find evidence that deflation was in fact more harmful during the Great Depression. In all specifications, the interaction term with the time-period dummy is statistically significant at least at the 10% level. Using IV the result reverses and there are no significant differences between the two samples. Meanwhile, the interaction term covering the entire sample is statistically significant and larger than the OLS estimate in three out of four cases.[25]

---

[24] The *F*-statistic is substantially larger at 34.4 and the estimates are more precise.

[25] In this specification, the rk F-statistic is 7.83 compared to a 5% critical value of 7.03 suggesting that we would reject the null hypothesis, that the IV bias is less than 10% of the OLS bias. Because we have more than one endogenous regressor, the critical values stem from Stock and Yogo (2005) and therefore the test comes with the caveat that the critical values from are formally justified only in the case of i.i.d. errors.

3.3. Evidence from modern data

The last solution to the measurement error problem is to use well-measured post-WWII data. We can base the analysis on the data collected by Jordà et al. (2016) and Knoll et al. (2016) for 17 industrialized economies.[26] The panel comprises annual CPI inflation and real per capita GDP growth. In addition, I calculate an HP-filtered output gap. As control variables, the data provide house prices, share prices and a systemic crisis indicator, and all regressions include country-time fixed-effects. Before turning to the empirical results, it is worth noting that the deflations were milder than during the 19th century US. Over all 17 countries and the entire time period, only 5% of the observations show a decline in the price level stronger than -3%. At the same time, the average deflation rate amounted to -1.0%. During the 19th century US, however, 60% of all measured deflations were stronger than -3% and the average deflation rate was -4.7%.

[Table 6 about here]

Table 6 presents the results using the deflation dummy as well as the deflation interaction term. Panel (A) shows that using GDP growth, none of the coefficients are statistically significant. For the output gap, however, there is a significant interaction term implying that a 1pp disinflation, if inflation is negative, is associated with a 1.2pp lower output gap. If we exclude euro area countries, the interaction term remains statistically significant (Panel B). In modern data, the CPI may systematically overestimate actual inflation because of neglected changes in quality, as emphasized by the Boskin Commission (1996). To take into account such a bias, Panel (C) provides estimates for a deflation threshold at 1%. For GDP growth the coefficients are still insignificant. For the output gap, the individual deflation dummy is significant, and also, the disinflation interaction term remains significant at least at the 10% level. Taking into account the level of inflation, dummy and interaction term, a decline in inflation from 0% to -1% is associated with 1.5pp lower real activity. The results based on modern data broadly confirm that deflation is associated with lower real economic activity, at least, against the backdrop of the more benign deflations during the post-WWII era. The main difference to the findings in Borio et al. (2015) stem from the fact that using an output gap as an independent variable yields a statistically significant relationship also for modern deflations.

4. Robustness and specification tests

This section discusses robustness and specification tests regarding the size of the bias using historical data, the IV-regressions, and modern data. Tables are included in Appendix C.

---

[26] In what follows, I focus on the post-WWII era. Results including the pre-WWII data are used in the next section for robustness tests.

4.1. Brackets for the bias

Instead of assessing the size of the bias using modern replications, we can apply statistical methods on historical data to calculate brackets for the true underlying coefficient and therefore the bias (see Hausman, 2001). Recall that the OLS estimate of (3) yields a lower bracket for the true coefficient because of the classification bias. We can estimate the reverse regression of equation (3):

$$1\{\tilde{\pi}_t < 0\} = -c\gamma + \gamma y_t - \gamma v_t, \tag{6}$$

with $\gamma = 1/\delta$. The OLS estimate of the slope coefficient ($\hat{\gamma}_{OLS}$) will be attenuated towards zero because y_t is by construction negatively correlated with the error term. Note that v_t includes ε_t from the original equation (1). Therefore, the inverse of the OLS estimate will be biased away from zero and $[\hat{\delta}_{OLS}, 1/\hat{\gamma}_{OLS}]$ yields a bracket for the true value of δ. How precisely this bracket can be estimated also depends on whether the real activity measure contains additional measurement error.

I performed reverse regressions for the period 1800-1945 using the four measures of real activity. The upper brackets are less precisely estimated than the lower brackets. For example, a 95% confidence interval for the upper bracket based on the reverse regression using GDP growth amounts to [-48.4, -18.9]. Meanwhile, the confidence interval amounts to [-4.5, -1.1] in the original regression. A conservative assessment of the possible misclassification factor therefore amounts to 0.2. This is qualitatively in line with the attenuation and classification biases implied by the modern replications. The results are similar for the other measures of real activity. Moreover, when using the wholesale price index, the brackets are in line with the brackets based on the CPI, suggesting relevant measurement error also in the WPI inflation rate.

4.2. IV-regressions

For the IV-regressions, I performed various robustness and specification tests using historical and modern US data. The IV-regressions based on the proxy from 1800-1890 are qualitatively robust when changing the deflation threshold to 1%. This robustness test takes into account that a quality-related systematic bias may also affect retrospective historical CPI estimates. We can also restrict the sample to moderate inflations and deflations between -5% and 5%. The coefficients become somewhat smaller in absolute size but, when using IV, remain statistically significant at least at the 10% level. I also examined deflationary periods where both, the CPI and the proxy, agree on whether we observe an inflationary or deflationary episode. If both measures give an independent signal whether prices were rising or falling, we have more confidence in the signal when both agree. Using this more naïve and less efficient approach, the results are only slightly less pronounced.

We can also perform various specification tests to check the validity and strength of the instrument. First, we can identify the coefficient in the reverse regression using the same set of instruments, a specification test proposed by Hahn and Hausman (2002). The results are basically identical. Second, we can use the CPI deflation dummy as an instrument for the proxy variable. Using IV, there is still negative association in three out of four cases. But the difference to OLS is less pronounced.[27]

For the IV-regressions, we assumed that deflation in terms of wholesale prices has no impact on real activity except through the common unobserved trend with CPI inflation. This is an exclusion restriction stating that a well-measured wholesale price deflation dummy should not be added to equation (1). Although we cannot test this exclusion restriction on historical data, we can examine the modern data from 1957-2015. If we are willing to assume that modern CPI data is essentially measured without error we can add a wholesale price deflation dummy to equation (1) and test whether it is statistically significant. Because we do not observe many deflations during this episode, I perform this test with an artificial threshold at 3%. This is somewhat lower than the average inflation rate and implies that half of the sample is classified as artificial deflations. Therefore, this is also a placebo test because the CPI deflation dummy should be insignificant in this case. Both, the artificial CPI deflation and WPI deflation dummies, are statistically insignificant. Testing the exclusion restriction without an artificial threshold, including CPI inflation and WPI inflation as continuous variables, yields qualitatively similar results.

To check whether the results also hold for other countries, I extended the analysis using panel data on 17 countries collected by Jordà et al. (2016) and Knoll et al. (2016). The data set includes annual data from 1870-2015.[28] Unfortunately, the data set lacks wholesale prices to instrument for the possibly error-ridden CPI. But, even if wholesale prices are added, it would be difficult to establish the validity of the instrument because we would have to check for every country whether the data sources overlap.[29] As an alternative, I construct an instrument based on broad money growth lagged by one period. Although this instrument should be correlated with CPI inflation and uncorrelated with measurement error in the CPI, it is also uncorrelated with supply shocks and therefore may overstate the actual correlation of deflation with real activity. Therefore, this specification is reported only as a robustness test.

Whether using OLS or IV, the results show no significant association between GDP growth and deflation. Using the output gap, however, the IV-regressions show a substantially stronger statistically significant association. The result remains similar if we exclude the US. If we additionally include an interaction term allowing for a different association after 1914, no significant difference emerges. However, this result should be discounted because the rk *F*-statistic is quite low at 4.5 compared to a

[27] Possibly, wholesale prices are better measured than consumer prices as suggested by Eichengreen et. al (2016). But also, this could stem from the fact that wholesale prices are more volatile and therefore the signal to noise ratio is higher for the same amount of measurement error.
[28] House prices for the US start only in 1890.
[29] For Switzerland, for example, the CPI for first half of the 19th century is identical to a WPI (see Studer and Schuppli, 2008).

5% critical value of 7.03.[30] Overall, this confirms the results based on modern data, that deflation is associated with a lower output gap, but, not necessarily with lower GDP growth.

4.3. Modern data

To test the robustness of the results based on modern panel data I study recent deflationary episodes for a panel of 15 Euro Area member countries. Today's central bankers aim to avoid potentially harmful deflations. Therefore, a typical deflationary episode may be relatively benign when central banks offset short-falls in aggregate demand but do not respond to beneficial supply shocks. If this is the case, reduced-form regressions based on modern data may suffer from the Lucas (1976) critique and are valid only under policy regimes that avoid harmful deflations. By contrast, the metal currency regimes of the 19th century were accompanied by regular deflationary periods that can be regarded as a necessary consequence of committing to a Gold Standard rule (see e.g. Bordo and Kydland, 1996).

A monetary union with a low average inflation rate is an interesting case to study because, if inflation is low on average, some member countries will likely experience falling prices, whereas for others, the general level of prices is rising. The member countries cannot use monetary policy to individually address deflationary pressures because the common central bank focuses on avoiding deflation in terms of the average.[31] The annual data cover 15 Euro Area member states, span the period from 2007 to 2015, and stem from OECD. The data include CPI inflation, an output gap, the unemployment rate, an estimate of the NAIRU, real house price changes as well as real share price changes. An additional advantage of this data set therefore is that we can examine the unemployment rate and a NAIRU-based unemployment gap as dependent variables. A disadvantage is, however, that the average deflation is even milder at only -0.6% and no deflationary episode showed a decline in the price level of more than -2%.

The modern Euro Area data give additional but substantially weaker evidence that deflation was associated with a lower output gap and higher unemployment. For the entire sample, there is no significant association between GDP growth and the deflation dummy. At least at the 10% significance level a link emerges for the output gap, the unemployment rate, and the unemployment gap. The disinflation interaction term is insignificant. The coefficient on inflation itself, however, is statistically significant suggesting that a disinflation inflation is associated with a lower output gap, higher

[30] The critical values stem from Stock and Yogo (2005) for two endogenous regressor. Note that they are not formally justified in the presence of HAC-robust standard errors.
[31] Figure C.1. in Appendix C shows that in the wake of the financial and Euro Area debt crises the Euro Area inflation rate declined to about 0% since 2013. This implied that 5 of the 15 member states experienced on average deflation in terms of the CPI. Meanwhile, because of the Euro Area debt crisis, fiscal policy was not available to stimulate demand.

unemployment rate, and higher unemployment gap. Increasing the deflation threshold to 1% does not materially alter this result.

5. Concluding remarks

This paper shows that estimating average real economic performance during deflations is hampered by measurement error in historical CPI data. Replications of deficiencies in 19th century CPI estimates suggest that those measurement issues are relevant and may explain some of the strikingly different time series properties of CPI inflation between the 19th century and the post-WWII era. Those deficiencies increase inflation volatility, reduce inflation persistence and attenuate the link between real economic activity and deflation. To estimate average growth during 19th century deflations, an IV-regression approach alleviates the errors-in-variables problem. I find that deflations were associated with lower real activity. The most surprising finding, perhaps, implies that the Great Depression was not significantly different from other deflationary episodes during the 19th century. The deflationary pressures during the 19th century were substantially stronger than what we find during the post-WWII period. This may explain the finding that the association between real activity and deflation became weaker in modern data and is limited to the output gap and unemployment.

Many empirical studies using 19th century data fail to uncover a significant link between real economic activity and deflation. A possible explanation is that 19th century deflations were benign, short-lived, or a by-product of beneficial advances in productivity. In addition, researchers find that during the 19th century prices and wages were quite flexible. To the extent that nominal rigidities are associated with a high persistence of inflation, the findings suggest that we may not only underestimate GDP growth during deflations, but also, the degree of nominal rigidities during the 19th century.

Nevertheless, this paper remains silent on whether deflation causes lower real activity or whether it is a consequence of falling aggregate demand. Therefore, it does not take a stand on whether deflation is harmful in because of a particular nominal rigidity or whether the negative association stems from more-regular negative aggregate demand relative to beneficial aggregate supply disturbances. Accurately estimating the reduced-form correlation between real activity and deflation, however, is a necessary condition for reliable structural analysis. Most estimation approaches to identify the impact of structural shocks will likely suffer from the errors-in-variables problem. Examining the impact of measurement error on structural analysis is beyond the scope of this paper but would be an interesting avenue for future research.

References

Adams, T. M. (1939), *Prices Paid by Farmers for Goods and Services and Received by Them for Farm Products, 1790-1871; Wages of Farm Labor, 1789-1937: A Preliminary Report*. University of Vermont and State Agricultural College, Vermont Agricultural Experiment Station.

Aigner, D. J. (1973), "Regression with a Binary Independent Variable Subject to Errors of Observation," *Journal of Econometrics*, 1(1), 49-59.

Allen, S. G., (1992), "Changes in the Cyclical Sensitivity of Wages in the United States, 1891-1987," *American Economic Review*, 82(1), 122-140.

Atkeson, A. and P. Kehoe (2004), "Deflation and Depression: Is There an Empirical Link?" *American Economic Review*, 94(2), 99-103.

Barsky, R. B. (1987), "The Fisher Hypothesis and the Forecastability and Persistence of Inflation," *Journal of Monetary Economics*, 19, 3-24.

Barsky, R. B. and J. B. DeLong (1991), "Forecasting Pre-World War I Inflation: The Fisher Effect and the Gold Standard," *The Quarterly Journal of Economics*, 106, 815-836.

Bayoumi, T. and B. Eichengreen (1996), "The Stability of the Gold Standard and the Evolution of the International Monetary Fund System," in T. Bayoumi, B. Eichengreen and M. P. Taylor (eds), *Modern Perspectives on the Gold Standard*, Cambridge: Cambridge University Press.

Benati, L. (2008), "Investigating Inflation Persistence across Monetary Regimes", *The Quarterly Journal of Economics*, 123(3), 1005-1060.

Bernanke, B. S. (2002), "Deflation: Making Sure 'It' Doesn't Happen Here," Remarks by Governor Ben S. Bernanke before the National Economists Club, Washington, D.C.

Bernholz, P. (2003), *Monetary Regimes and Inflation—History, Economics and Political Relationships*, Cheltenham: E. Elgar.

BLS (1923), *Retail Prices 1913 to December, 1921*, Bulletin No. 315, U.S. Bureau of Labor Statistics.

Bordo, M. D. and A. Filardo (2005), "Deflation and Monetary Policy in a Historical Perspective: Remembering the Past or Being Condemned to Repeat it?" *Economic Policy*, 20, 799-844.

Bordo, M. D. and F. E. Kydland (1996), "The Gold Standard as Commitment Mechanism," in T. Bayoumi, B. Eichengreen and M. P. Taylor (eds), *Modern Perspectives on the Gold Standard*, Cambridge: Cambridge University Press.

Bordo, M. D. and A. Redish (2004), "Is Deflation Depressing? Evidence from the Classical Gold Standard," in R. Burdekin and P. Siklos (eds), *Deflation: Current and Historical Perspectives*, Cambridge: Cambridge University Press.

Borio, C., M. Erdem, A. Filardo, and B. Hofmann (2015): "The Costs of Deflations: A Historical Perspective," *BIS Quarterly Review*.

Boskin Commission (1996), *Toward a More Accurate Measure of the Cost Of Living: Final Report to the Senate Finance Committee from the Advisory Commission to Study the Consumer Price Index*, retrieved from: http://www.ssa.gov/history/reports/boskinrpt.html.

Cagan, P. (1975), "Changes in the Recession Behavior of Wholesale Prices in the 1920s and Post-World War II," in *Explorations in Economic Research*, NBER Chapters, 2(1), 54-104.

David, P. A., and P. Solar (1977), "A Bicentenary Contribution to the History of the Cost of Living in America." In P. Uselding (ed), *Research in Economic History*, vol. 2, Greenwich: JAI Press.

Davis, J. H. (2004), "An Annual Index of U.S. Industrial Production, 1790-1915," *The Quarterly Journal of Economics*, 119, 1177-1215.

Davis, J. H., C. Hanes, and P. W. Rhode (2009), "Harvests and the Business Cycle in Nineteenth-Century America", *The Quarterly Journal of Economics*, 124(4), 1675-1727.

Eichengreen, B., D. Park, and K. Shin (2016), "Deflation in Asia: Should the Dangers Be Dismissed?" Asian Development Bank Economics Working Paper Series, No. 490.

Fabricant, S. (1940), *The Output of Manufacturing Industries, 1899-1937*, National Bureau of Economic Research.

Friedman, M. (1969), "The Optimum Quantity of Money," in *The Optimum Quantity of Money and Other Essays*, Chicago: Aldine.

1–50.Gordon, R. J (1980), "A Consistent Characterization of a Near-Century of Price Behavior," *American Economic Review*, 70(2), 243-249.

Gordon, R. J. (2015), *The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War.* Princeton: Princeton University Press.

Hahn, J. and J. Hausman (2002), "A New Specification Test for the Validity of Instrumental Variables," *Econometrica*, 70, 163-189.

Hanes, C. (1998), "Consistent Wholesale Price Series for the United States, 1860-1990," in T. J. O. Dick (ed), *Business Cycles since 1820: New International Perspectives from Historical Evidence*. Cheltenham: E. Elgar.

Hanes, C. (2006), "Wholesale Price indexes, by Commodity Group: 1749-1890 [Warren and Pearson]," in S. B. Carter, S. S. Gartner, M. R. Haines, A. L. Olmstead, R. Sutch, and G. Wright (eds), *Historical Statistics of the United States, Earliest Times to the Present: Millennial Edition*, New York: Cambridge University Press.

Hansen, B. E. (1999), "The Grid Bootstrap and the Autoregressive Model," *The Review of Economics and Statistics*, 81(4), 594-607.

Hausman, J. (2001), "Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left," *Journal of Economic Perspectives*, 15(4), 57-67.

Hoover, E. D. (1960), "Retail Prices after 1850," in *Trends in the American Economy in the Nineteenth Century*, 141-86. Studies in Income and Wealth, vol. 24, National Bureau of Economic Research. Princeton: Princeton University Press.

Jalil, A. J. (2015), "A New History of Banking Panics in the United States, 1825-1929: Construction and Implications," *American Economic Journal: Macroeconomics*, 7(3), 295-330.

Johnston, L. and S. H. Williamson (2016), *What Was the U.S. GDP Then?*, MeasuringWorth, retrieved from http://www.measuringworth.org/usgdp/.

Jordà, O., M. Schularick, and A. M. Taylor (2016), "Macrofinancial History and the New Business Cycle Facts," in M. Eichenbaum and J. A. Parker (eds), *NBER Macroeconomics Annual 2016*, 31, Chicago: University of Chicago Press.

Kleibergen, F. and R. Paap (2006), "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics*, 133, 97-126.

Klein, B. (1975), "Our New Monetary Standard: The Measurement and Effects of Price Uncertainty, 1880-1973," *Economic Inquiry*, 13, 461-484.

Knoll, K., M. Schularick, and T. Steger (2016), "No Price Like Home: Global House Prices, 1870-2012," *American Economic Review*, forthcoming.

Lebergott, S. (1964), *Manpower in Economic Growth: The American Record since 1800*. New York: McGraw-Hill.

Lucas, R. E. (1976): “Econometric Policy Evaluation: A Critique,” *Carnegie-Rochester Conference Series on Public Policy*, 1, 19-46.

Long, C. D. (1960), *Wages and Earnings in the United States, 1860-1890*. National Bureau of Economic Research. Princeton: Princeton University Press.

Montiel Olea, J. L. and C. E. Pflueger (2013), “A Robust Test for Weak Instruments,” *Journal of Business and Economic Statistics*, 31, 358-369.

Ng, S. and P. Perron (1995), “Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag,” *Journal of the American Statistical Association*, 90(429), 268-281.

Officer, L. H. and S. H. Williamson (2016a), *The Annual Consumer Price Index for the United States, 1774-2015*, MeasuringWorth, retrieved from: http://www.measuringworth.com/uscpi/.

Officer, L. H. and S. H. Williamson (2016b), *Annual Wages in the United States, 1774-Present*, MeasuringWorth, retrieved from: http://www.measuringworth.com/uswages/.

Officer, L. H. (2014), *What Was the Consumer Price Index Then? A Data Study*, MeasuringWorth, retrieved from: http://www.measuringworth.com/docs/cpistudyrev.pdf.

Rees, A. (1961), *Real Wages in Manufacturing, 1890-1914*, National Bureau of Economic Research, Princeton: Princeton University Press.

Romer, C. (1986a), “Spurious Volatility in Historical Unemployment Data,” *Journal of Political Economy*, 94, 1-37.

Romer, C. (1986b), “Is the Stabilization of the Postwar Economy a Figment of the Data?” *American Economic Review*, 76, 314-334.

Sachs, J. (1980), “The Changing Cyclical Behavior of Wages and Prices: 1890-1976”, *American Economic Review*, 70(1), 78-90.

Sargent, T. J. (1973), “Interest Rates and Prices in the Long Run: A Study of the Gibson Paradox,” *Journal of Money, Credit and Banking*, 5, 385-449.

Schwert, G. W. (1989), “Tests for Unit Roots: A Monte Carlo Investigation,” *Journal of Business & Economic Statistics*, 7, 147-160

Shiller, R. J. and J. J. Siegel (1977), “The Gibson Paradox and Historical Movements in Real Interest Rates,” *Journal of Political Economy*, 85, 891-907.

Shoemaker, O. (2014), *Variance Estimates for Price Changes in the Consumer Price Index: January-December 2014*, U.S. Bureau of Labour Statistics, retrieved from: http://www.bls.gov/cpi/cpivar2014.pdf.

Staudenmayer, J. and J. P. Buonaccorsi (2005), “Measurement Error in Linear Autoregressive Models”, *Journal of the American Statistical Association*, 100(471), 841-852.

Stock J. H. and M. Yogo (2005), “Testing for Weak Instruments in Linear IV Regression,” in: D. W. K. Andrews and J. H. Stock (eds), *Identification and Inference for Econometric Models*. New York: Cambridge University Press.

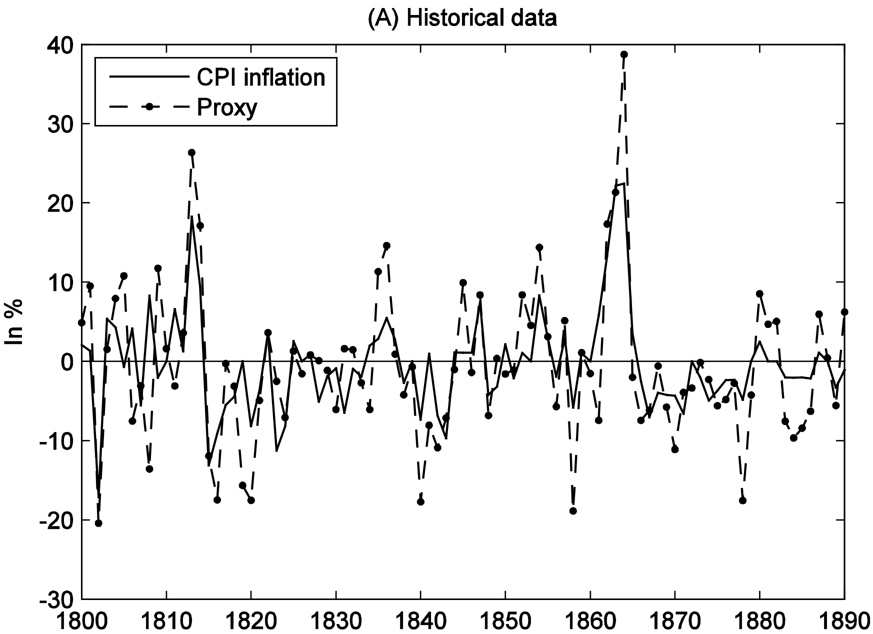
Studer R. and P. Schuppli (2008), “Deflating Swiss Prices over the Past Five Centuries,” *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 41(3), 137-156.

Warren, G. and F. Pearson (1933), *Prices*, New York: George Wiley and Sons.

Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, Cambridge: The MIT Press.

Figures and tables

Figure 1. Historical and modern proxies based on wholesale prices



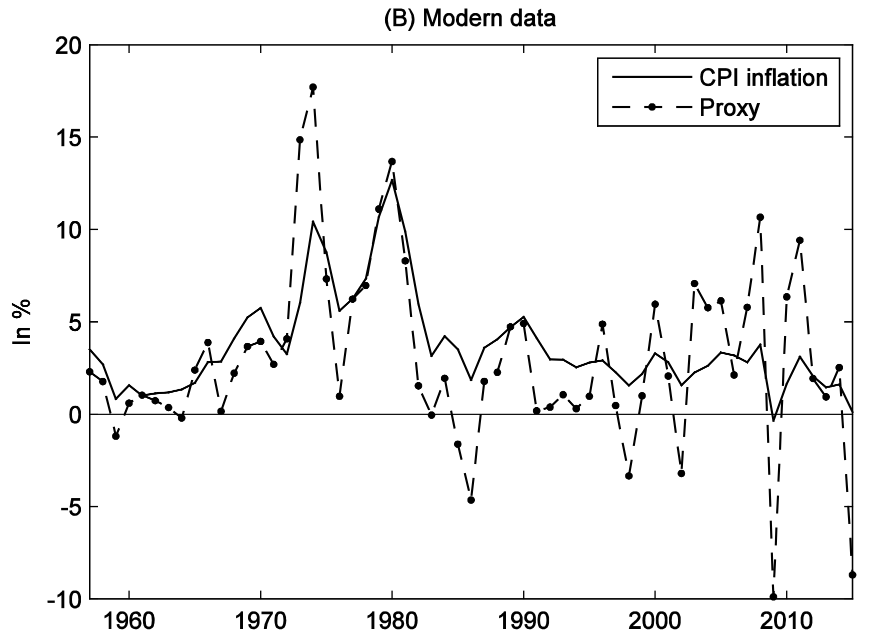


Table 1. Selected methodological deficiencies in retrospective CPI estimates

Deficiency	Source	Time span	Comments
Wholesale prices	David and Solar (1977)	1774-1851	Wholesale prices for Philadelphia approximate retail prices before 1800
Geographical coverage	David and Solar (1977)	1774-1851	Wholesale prices for Philadelphia (before 1800) and retail prices paid by Vermont farmers (until 1851)
Sample size	Hoover (1960)	1851-1860	Weeks Report shows many missing observations and small number of individual price quotes
	Long (1960)	1880-1890	Little information on retail prices for the entire decade after the Weeks Report ends
Linear interpolation	Long (1960)	1880-1890	Several items interpolated linearly over the decade (particularly rent)
Reproduction cost index	Lebergott (1964)	1860-1880	Rent approximated by prices of construction materials and wages of low-skilled workers
Few services	Lebergott (1964)	1860-1880	Almost no services included

Note: The time span represents the segment used in the composite CPI by Officer and Williamson (2016a).

Table 2. Descriptive statistics and replications

(A) CPI Inflation

	Standard deviation	Persistence	Share deflation	Attenuation factor	Misclassification factor
1800-1913	5.7	0.43	0.46	-	-
1914-1956	6.6	0.48	0.28	-	-
1957-2015	2.6	0.85	0.02	-	-

(B) Individual replications 1957-2015

	Standard deviation	Persistence	Share deflation	Attenuation factor	Misclassification factor
(1) 2013 weights	2.6	0.87	0.00	1.0	1.0
(2) 1869 weights	2.8	0.70	0.03	0.9	0.5
(3) Philadelphia only	2.6	0.80	0.03	1.0	0.5
(4) PPI for nondurables	2.8	0.78	0.02	0.9	1.0
(5) PPI for durables	2.6	0.86	0.00	1.0	1.0
(6) Linear interpolation shelter	2.4	0.84	0.03	1.2	0.5
(7) Reproduction cost index	2.5	0.80	0.03	1.1	0.5
(8) Missing services	3.0	0.66	0.07	0.8	0.3
(9) Sampling error	2.7	0.79	0.03	0.9	0.4

(C) Combined replications 1957-2015

	Standard deviation	Persistence	Share deflation	Attenuation factor	Misclassification factor
Combination at 2013 weights	5.3	0.28	0.21	0.2	0.1
Combination at 1869 weights	5.3	0.42	0.24	0.2	0.1

Note: The table shows descriptive statistics for actual CPI inflation (Panel A) and replications of methodological deficiencies in 19th century CPI inflation measures (Panel B). Panel (C) combines the most serious deficiencies (replications 4, 6, 7, 8, 9) using different expenditure weights. The persistence is the sum of autoregressive coefficients, where the number of lags is determined following Ng and Perron (1995). The maximum number of lags in the autoregressive model is determined according to a rule of thumb (see Schwert, 1989). The final lag length is determined by iteratively reducing the lag length as long as the *t*-statistic of the last autoregressive term is larger than 1.6. The third column shows the share of years with negative inflation. The attenuation factor is the true variance of inflation divided by the variance of the replication. The misclassification factor is calculated as unity minus the share of misclassified inflations and deflations. Descriptive statistics involving sampling error are means of 10,000 simulations, adding draws from an i.i.d. normal distribution with sampling error variance scaled by the relative number of observations in modern and historical price data.

Table 3. Real economic performance during deflations

	(1) ΔGDP_t	(2) ΔGDP_t	(3) ΔIP_t	(4) ΔIP_t	(5) $\widehat{GDP}_t$	(6) $\widehat{GDP}_t$	(7) $\widehat{IP}_t$	(8) $\widehat{IP}_t$
				(A) No control variables (1800-1890)				
$1\{\tilde{\pi}_t < 0\}$	-1.26	-3.60*	-1.69	-8.19**	-2.20**	-6.16***	-2.63	-9.03***
	(0.65)	(1.45)	(1.45)	(3.00)	(0.72)	(1.54)	(1.35)	(2.61)
Estimator	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Observations	91	91	91	91	91	91	91	91
				(B) Additional control variables (1800-1890)				
$1\{\tilde{\pi}_t < 0\}$	-1.53*	-4.48**	-2.00	-9.69**	-2.42**	-7.25***	-3.01*	-10.69***
	(0.66)	(1.53)	(1.57)	(3.44)	(0.76)	(1.80)	(1.37)	(3.02)
Estimator	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Observations	88	88	88	88	88	88	88	88

Note: The table shows regressions of the form $y_t = c + \delta 1\{\tilde{\pi}_t < 0\} + \theta' controls_t + \nu_t$. The indicator function $1\{\tilde{\pi}_t < 0\}$ assumes unity if CPI inflation is negative and zero otherwise. The IV-regressions use a deflation dummy based on the proxy variable as an instrument. Control variables include equity price inflation and dummies for major banking crises (Jalil, 2015). Coefficients on constants and controls are not shown. Columns (1)-(4) give the results for GDP and industrial production in growth rates, and columns (5)-(8) give the results for the HP-filtered measures. HAC-robust standard errors are given in parentheses. Coefficients with superscripts ***/**/* are statistically significant at the 1%/5%/10% level.

Table 4. Differences during the Great Depression

	(1) ΔGDP_t	(2) ΔGDP_t	(3) ΔIP_t	(4) ΔIP_t	(5) $\widehat{GDP}_t$	(6) $\widehat{GDP}_t$	(7) $\widehat{IP}_t$	(8) $\widehat{IP}_t$
				(A) WPI as instrument (1800-1945)				
$1\{\tilde{\pi}_t < 0\}$	-2.90**	-7.25***	-3.35*	-14.76***	-3.86***	-6.60***	-6.40***	-13.77***
	(0.87)	(1.74)	(1.68)	(3.66)	(0.99)	(1.86)	(1.69)	(3.18)
Estimator	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Observations	143	143	143	143	143	143	143	143
				(B) WPI as instrument with interaction term (1800-1945)				
$1\{\tilde{\pi}_t < 0\}$	-1.15	-5.74**	-1.30	-13.64***	-2.70***	-7.38***	-3.54**	-12.51***
	(0.76)	(1.82)	(1.48)	(4.02)	(0.71)	(1.80)	(1.26)	(3.18)
$1\{\tilde{\pi}_t < 0\} \times$	-8.00***	-5.20	-10.98*	-7.28	-5.47	3.18	-13.66*	-4.93
$1\{t > 1914\}$	(2.27)	(3.89)	(5.32)	(9.27)	(3.88)	(5.77)	(6.40)	(9.03)
Estimator	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Observations	143	143	143	143	143	143	143	143

Note: The table shows regressions of the form $y_t = c + \delta 1\{\tilde{\pi}_t < 0\} + \theta' controls_t + \nu_t$. The indicator function $1\{\tilde{\pi}_t < 0\}$ assumes unity if CPI inflation is negative and zero otherwise. The IV-regressions include deflation dummies based on the wholesale price index as instruments. All regressions include as control variables equity price inflation and dummies for major banking crises (Jalil, 2015). Coefficients on constants and controls are not shown. Columns (1)-(4) give the results for GDP and industrial production in growth rates, and columns (5)-(8) give the results for the HP-filtered measures. HAC-robust standard errors are given in parentheses. Coefficients with superscripts ***/**/* are statistically significant at the 1%/5%/10% level.

Table 5. Real economic performance during disinflations

	(1) ΔGDP_t	(2) ΔGDP_t	(3) ΔIP_t	(4) ΔIP_t	(5) $\widehat{GDP}_t$	(6) $\widehat{GDP}_t$	(7) $\widehat{IP}_t$	(8) $\widehat{IP}_t$
(A) Proxy as instrument (1800-1890)								
$\tilde{\pi}_t \times 1\{\tilde{\pi}_t < 0\}$	0.44**	1.93*	0.70*	5.57*	-0.17	1.88	-0.25	4.26
	(0.16)	(0.78)	(0.33)	(2.49)	(0.20)	(1.13)	(0.39)	(2.29)
Estimator	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Observations	88	88	88	88	88	88	88	88
(B) WPI as instrument (1800-1945)								
$\tilde{\pi}_t \times 1\{\tilde{\pi}_t < 0\}$	0.46*	1.81**	1.19*	6.69***	0.23	1.12*	0.77	4.46***
	(0.21)	(0.60)	(0.51)	(1.41)	(0.29)	(0.48)	(0.67)	(0.97)
Estimator	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Observations	143	143	143	143	143	143	143	143
(C) WPI as instrument with interaction terms (1800-1945)								
$\tilde{\pi}_t \times 1\{\tilde{\pi}_t < 0\}$	0.22	1.71**	0.65	5.89**	-0.05	1.34	-0.03	3.60**
	(0.19)	(0.61)	(0.38)	(1.77)	(0.25)	(0.71)	(0.44)	(1.31)
$\tilde{\pi}_t \times 1\{\tilde{\pi}_t < 0\} \times$	1.00**	0.28	1.77**	1.06	1.01*	-0.18	2.91***	1.23
$1\{t > 1914\}$	(0.32)	(0.56)	(0.59)	(1.31)	(0.42)	(0.48)	(0.48)	(0.88)
Estimator	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Observations	143	143	143	143	143	143	143	143

Note: The table shows regressions of the form $y_t = c + \beta_1 \tilde{\pi}_t + \beta_2 \tilde{\pi}_t \times 1\{\tilde{\pi}_t < 0\} + \beta_3 1\{\tilde{\pi}_t < 0\} + \theta' controls_t + \nu_t$. The indicator function $1\{\tilde{\pi}_t < 0\}$ assumes unity if CPI inflation is negative and zero otherwise. As instruments for the error-ridden interaction terms, the IV-regressions include the corresponding terms based on the proxy variable or the wholesale price index. All regressions include as control variables equity price inflation and dummies for major banking crises (Jalil, 2015). Coefficients on constants and controls are not shown. Columns (1)-(4) give the results for GDP and industrial production in growth rates, and columns (5)-(8) give the results for the HP-filtered measures. HAC-robust standard errors are given in parentheses. Coefficients with superscripts ***/**/* are statistically significant at the 1%/5%/10% level.

Table 6. Real economic performance during modern deflations

	(1) ΔGDP_{it}	(2) ΔGDP_{it}	(3) $\widehat{GDP}_{it}$	(4) $\widehat{GDP}_{it}$
(A) 17 countries (1950-2015)				
$1\{\pi_{it} < 0\}$	-0.25	-0.52	-0.82	0.28
	(0.37)	(0.39)	(0.42)	(0.39)
π_{it}		-0.04		0.11**
		(0.03)		(0.03)
$\pi_{it} \times 1\{\pi_{it} < 0\}$		-0.24		1.13***
		(0.20)		(0.30)
R-squared	0.58	0.58	0.49	0.52
Observations	986	986	986	986
(B) Excluding Euro Area countries (1950-2015)				
$1\{\pi_{it} < 0\}$	0.64	0.02	-1.57*	-0.28
	(0.76)	(0.82)	(0.76)	(0.64)
π_{it}		-0.10		0.09*
		(0.06)		(0.05)
$\pi_{it} \times 1\{\pi_{it} < 0\}$		-0.30		1.01***
		(0.24)		(0.23)
R-squared	0.44	0.45	0.45	0.48
Observations	434	434	434	434
(C) Higher threshold (1950-2015)				
$1\{\pi_{it} < 1\}$	-0.14	-0.22	-0.71**	-0.69**
	(0.19)	(0.22)	(0.22)	(0.24)
π_{it}		-0.04		0.10**
		(0.04)		(0.03)
$\pi_{it} \times 1\{\pi_{it} < 1\}$		-0.02		0.66*
		(0.20)		(0.30)
R-squared	0.58	0.58	0.50	0.52
Observations	986	986	986	986

Note: The table shows regressions of the form $y_{it} = c_i + d_t + \beta_1 \pi_{it} + \beta_2 \pi_{it} \times 1\{\pi_{it} < 0\} + \delta 1\{\pi_{it} < 0\} + \theta' controls_{it} + \nu_{it}$. The indicator function $1\{\pi_{it} < 0\}$ assumes unity if CPI inflation is negative in country i and zero otherwise. All regressions include country-time fixed-effects. Other control variables include real house price as well as equity price inflation, and a systemic crisis dummy. Coefficients on constant and controls are not shown. The dependent variables are GDP growth and an HP-filtered output gap. HAC-robust standard errors are given in parentheses. Coefficients with superscripts ***/**/* are statistically significant at the 1%/5%/10% level.

Appendix A. Data sources and descriptive statistics

Table A.1. Data and sources

Name	Time span	Source	Identifier	Comments
Composite price and wage indices				
CPI	1774-2015	MW		Officer and Williamson (2016a)
WPI	1749-1890	HSUS	Cc113	Warren and Pearson (1933)
	1860-1990	HSUS	Cc125	Hanes (1998)
	1913-2015	FRED	PPIACO	
Nominal wages	1774-2015	MW		Unskilled workers; Officer and Williamson (2016b)
BLS CPI data				
All item U.S. city average	1913-2015	BLS	0000SA0	Prefix CUUR applies to all BLS CPI identifiers
Philadelphia-Wilmington-Atlantic City	1914-2015	BLS	A102SA0	
All items less shelter	1935-2015	BLS	0000SA0L2	
Shelter	1952-2015	BLS	0000SAH1	
Nondurables	1935-2015	BLS	0000SAN	
Durables	1935-2015	BLS	0000SAD	
Services	1935-2015	BLS	0000SAS	Monthly data as of 1956
BLS PPI data				
Nondurables	1947-2015	FRED	DUR0120	Prefix WPU applies to all FRED PPI identifiers
Durables	1947-2015	FRED	DUR0110	
Processed foods and feeds	1947-2015	FRED	02	
Apparel	1947-2015	FRED	0381	
Fuels and related products and power	1926-2015	FRED	05	
Textile house furnishings	1947-2015	FRED	0382	
Construction Materials	1947-2015	FRED	SI012011	
Warren and Pearson WPI data				
Foods	1798-1890	HSUS	Cc115	Warren and Pearson (1933)
Textile products	1798-1890	HSUS	Cc117	Warren and Pearson (1933)
Fuel and lighting	1798-1890	HSUS	Cc118	Warren and Pearson (1933)
House furnishing goods	1840-1890	HSUS	Cc122	Warren and Pearson (1933)

Continued on next page

Table A.1. Data and sources – *continued from previous page*

Name	Time span	Source	Identifier	Comments
Composite real activity				
Per capita real GDP	1790-2015	MW		Johnston and Williamson (2016)
Industrial production	1790-1915			Davis (2004)
	1899-1937	HSUS	Dd495	Fabricant (1940)
	1919-2015	FRED	INDPRO	
GDP output gap	1790-2015			Own calculations; smoothing parameter set to 100.
Industrial production output gap	1790-2015			Own calculations; smoothing parameter set to 100.
Historical control variables				
Banking crises	1833-1834, 1837-1839, 1857, 1873, 1893, 1907	Jalil (2015)		
Equity prices	1802-1870	HSUS	Cj797	Index of common stocks
	1870-2015	JST		
Euro area data				
Annual data for EA, AT, BE, EE, FI, FR, DE, GR, IE, IT, LU, NL, PT, SK, SI, ES				
CPI inflation	2006-2015	OECD		
GDP growth	2006-2015	OECD		
Output gap	2006-2015	OECD		No data for Luxembourg
Share prices	2006-2015	OECD		Deflated by CPI inflation
House prices	2006-2015	OECD		In real terms
Unemployment rate	2006-2015	OECD		
Unemployment gap	2006-2015	OECD		Own calculations; Unemployment rate - NAIRU
Historical data for other countries				
Annual data for AU, BE, CA, DK, FI, FR, DE, IT, JP, NL, NO, PT, SE, ES, CH, UK, US, retrieved from macrohistory.net on 5 November 2016.				
CPI inflation	1870-2015	JST		
GDP growth	1870-2015	JST		
Output gap	1870-2015	JST		Own calculations; smoothing parameter set to 100.
Share prices	1870-2015	JST		Deflated by CPI inflation
House prices	1870-2015	JST		Deflated by CPI inflation. US data start in 1890.
Systemic crisis	1870-2015	JST		Dummy variable

Note: All composite series spliced using the most recent series available. MW: MeasuringWorth; HSUS: Historical Statistics of the United States; FRED: Federal Reserve Bank of St. Louis Economic Data; BLS: U.S. Bureau of Labor Statistics; JST: Jordà et al. (2016) and Knoll et al. (2016).

Table A.2. Expenditure shares

Commodity group	1869	1940	2013
Nondurable and semi-durable goods	**67.6**	**45.4**	**22.6**
Food, alcohol for off-premises consumption	44.3	22.3	7.6
Tobacco, printed material, heating/lighting fuel	7.6	5.9	1.6
Clothing and footwear	9.9	10.1	3.1
Dry goods for making clothing at home	5.0	0.0	0.0
House furnishings, toys, games, sports equipment	0.8	0.8	1.2
Not invented yet	0.0	6.3	9.1
Durable goods	**9.1**	**11.5**	**10.9**
Furniture, floor coverings, house furnishings	4.5	2.8	1.4
Glassware, tableware	0.9	0.7	0.4
Sporting equipment, guns, ammunition	0.0	0.3	0.5
Books, musical instruments, luggage	0.9	0.8	0.6
Jewelry and watches	1.5	0.6	0.7
Horse-drawn vehicles	1.3	0.0	0.0
Not invented yet	0.0	6.3	7.3
Services	**23.3**	**43.1**	**66.5**
Rent	18.0*	13.2	15.5
Food services and accommodation		6.2	6.2
Contributions		1.5	2.7
Not invented or not purchased		22.2	42.1

Note: Based on Gordon (2015) but adjusted so that the subcomponents sum to the aggregate and the aggregates to 100. * Expenditure share for rent in 1869 based on David and Solar (1977).

Table A.3. Descriptive statistics according to different monetary regimes

	Sample	Mean	Standard deviation	Persistence	Share deflation
Consumer prices	1800-1833	0.2	4.3	0.16	0.32
	1834-1861	0.7	9.4	0.56	0.71
	1862-1878	-0.0	2.0	0.57	0.37
	1879-1913	2.4	6.7	0.48	0.29
	1914-1956	3.6	2.6	0.85	0.02
	1957-2015	0.2	4.3	0.16	0.32
Wholesale prices	1800-1833	-0.8	8.6	-0.08	0.53
	1834-1861	-0.2	7.9	0.28	0.43
	1862-1878	0.1	13.5	0.43	0.76
	1879-1913	0.3	5.3	0.34	0.40
	1914-1956	2.1	11.8	0.20	0.36
	1957-2015	3.1	4.5	0.46	0.15

Note: Descriptive statistics on samples according to different monetary regimes before WWII (see Bordo and Kydland, 1996): Bimetallism, de facto Gold Standard after Coinage Act, Greenback period, Gold Standard, the period including WWI & II. For comparability with the sample used in the main text, the post-WWII sample starts in 1957.

Appendix B. Number of observations and sampling error historical price data

Unfortunately, we do not observe the sampling standard error for retrospective CPI estimates. We can investigate, however, how many price quotes were used to calculate a typical historical CPI and how much the modern sampling standard error increases if we would estimate a modern CPI based on a smaller number of individual price quotes. This requires, first, an estimate of the sampling standard error of the modern CPI inflation rate; second, an assumption on how the sampling standard error depends on the number of observations; and third, the number of individual price quote observations underlying retrospective estimates of historical CPI inflation. This appendix discusses the sources of the three measures.

The sampling standard error for modern CPI inflation is published by the BLS (see Shoemaker, 2014). Because of the large number of observations, sampling error is a minor issue. Currently, the BLS collects more than 80,000 price quotes each month to calculate the CPI inflation rate. Therefore, the annual average inflation rate is based on more than 1,000,000 price quote observations.[32] Against the backdrop of such a large sample, we expect that today's CPI inflation rate is precisely estimated. Indeed, for a typical 12-month inflation rate in 2014, the sampling standard error amounts to 0.07 percent, and a 95% confidence interval around a 1% 12-month inflation rate amounts to [0.9, 1.1] percent. In what follows, I use the sampling standard error of the 12-month inflation rate to approximate the sampling standard error of the annual average inflation rate. Simulations, available upon request, indicate that this is a reasonable and inconsequential approximation.

We can derive a relationship between the sampling standard error and the number of observations based on two simplifying assumptions. First, assume that the aggregate CPI inflation rate is the unweighted average of the individual price changes and second, that those individual price changes are i.i.d. with finite variance s^2. Then, a central limit theorem applies according to which the CPI inflation rate converges to a normal distribution, with the mean equal to the true inflation rate and a sampling variance $\sigma^2 = s^2/N$. Because the BLS publishes an estimate of the sampling standard error (σ) as well as the number of observations (N), this formula allows one to back out the variance of individual price changes (s^2). Given this variance, we can then examine how the sampling standard error changes when reducing the number of individual price quotes.

We have to examine, however, whether the simplifying assumptions reasonably approximate the more complicated methodology that is used to construct the CPI. A typical monthly inflation rate in

[32] In what follows, I assume that if the BLS records a price quote it also observes the price change. Therefore, the actual number of observations used in calculating the inflation rate may be lower if new products are introduced. This is a conservative assumption because missing data is much more likely to affect the data collection for the 19th century than today's professionally organized survey schemes.

2014 is based on a sample of approximately 87,000 price quotes, and the sampling standard error amounts to 0.04 percent (see Shoemaker, 2014). The formula would predict that the standard error for the Northeast region, which is based on a sample of approximately 18,400 price quotes, should amount to 0.09 percent. This is smaller than but very close to the value reported by the BLS (0.10 percent). The formula predicts that the U.S. city food index, based on 34,800 observations, has a sampling standard error of 0.06 percent, again smaller than but very close to the reported value of 0.07 percent.

I now turn to the question of how many monthly price quotes underlie annual estimates of CPI inflation for the 19th and early 20th century. By 1921, the number of price quotes was already substantial, and thus, sampling error is not a major issue. A BLS bulletin from 1923 allows a lower bound to be gauged for the number of monthly price quotes collected in a year (BLS, 1923). By 1921, food prices were collected in 51 cities and for 28 items. The number of price quotes varied with the size of the cities from 10-15 (smaller cities) to 20-30 (larger cities). Assuming that, on average, 20 price quotes were collected each month for each item, the sample size amounted to 342,000 monthly price quote observations each year. For most other items, there is no reliable information on how many price quotes were collected. We know, however, that by 1919, the number of items for which price quotes were collected was also substantial (BLS, 1941). The BLS collected prices for the following commodity groups (number of items in parentheses): clothing (65), fuel and lighting (6), rent (1), house furnishings (24), and miscellaneous goods and services (39). Officer (2014) reports that all items other than food were collected in 32 urban areas. Few prices, however, were collected on a monthly basis, because the CPI was published only for selected months of the year. If I assume that only one price quote was collected each quarter for each item in each of the 32 urban areas, this yields another 17,000 monthly price quotes each year. Thus, by 1921, we obtain an estimate of the number of price quotes underlying the annual CPI inflation rate of 360,000 monthly observations.

Retrospective estimates for the 19th century are based on a substantially lower number of observations. One particular segment of the Officer and Williamson (2016a) composite CPI is constructed by Hoover (1960). From her detailed description of the underlying data set, we can derive a range of the number of price quotes available for a typical year. The index is largely based on data from the comprehensive Weeks Report. Hoover (1960) notes on p. 146 that "This is by far the most extensive compilation of retail prices available for the nineteenth century." The survey indeed covers an impressive number of cities and items.[33] The main difference between Hoover's index and a modern

[33] There is evidence that the information from Hoover (1960) gives us an upper bound to the number of observations used to compute retrospective CPIs for other segments. Other researchers have discarded almost half of the price quotes from the Weeks Report because prices for June 1 are not representative of the entire annual average and because price series were not continuously reported (see Officer, 2014). Long (1960) notes that the retail price data for the period 1880-1890, when the Weeks Report ended and before the BLS started to collect monthly data on food items, is even thinner. Moreover, before 1851, the only retail price data on an annual basis stem from the often used prices paid by Vermont farmers (Adams, 1939).

CPI is that the number of price quote observations available from the Weeks Report is substantially smaller.

Hoover (1960) documents that data for the Weeks Report was collected from one or two respondents in more than 40 cities. They were asked to retrospectively provide average annual prices for the years 1851-1880. If no average price could be provided, they could instead report the price of 1 June. The 60 items covered a large number of commodity groups, including food, clothing, rent, fuel and light, and other goods.[34] If we assume that the annual average price reported is equivalent to 12 monthly observations and all respondents reported on all items, the maximum number of price quote observations in a typical year from the Weeks Report would exceed 75,000 (see Table B.1).

Although this is substantially lower than what the BLS collected in 1921, it still overestimates the available information because the survey was far from complete. Less than one-third of all respondents reported prices for all items. Moreover, few responses covered all 30 years, with more observations available towards the end of the sample period. Finally, only half of the price quotes were reported as an average price, and as Hoover speculates, many of the average prices reported for the early sample were guesses rather than based on actual monthly information (note that this was a retrospective survey). Table 3 therefore gives the number of price quotes in a typical year based on various assumptions on how many observations were missing.

Table B.1. Estimated number of price quotes underlying 19th century CPI estimates

	No missing	Optimistic	Pessimistic
Cities	43	43	43
Respondents	2	1.5	1.5
Items reported	74	74/3	74/4
Implied monthly obs.	12	12/2+1/2	12/4+3/4
Fraction of years	1	0.8	0.5
Number of price quotes	76,368	8,273	2,237

Note: Estimated number of price quotes in a typical year underlying the CPI constructed by Hoover (1960) if there are no missing observations, for an optimistic scenario on the number of missing observations and for a pessimistic scenario. The total number of observations amounts to the product of the individual elements.

Under an optimistic scenario, the number of implied monthly observations amounts to just over 8,000 each year.[35] This scenario reflects the situation towards the end of the Weeks Report. I assume

[34] To bridge gaps in the Weeks Report, Hoover added price data from less-representative sources for fruit, shoe repairs and physician fees. Fruit prices were estimated from wholesale prices for Philadelphia, whereas shoe repairs and physician fees stem from Adams' (1939) prices paid by Vermont farmers. Moreover, she collected prices for newspapers from the library of congress. To estimate the number of observations underlying her retrospective CPI, I treat those four items as being based on the same number of observations as the items covered by the Weeks Report.

[35] To put those numbers in some perspective, the number of price quotes collected increased by a factor of 125 between 1850 and 2014. At the same time, the resident population in the US increased only by a factor of 14, from 23 million to 319 million.

that, on average, 1.5 surveyed individuals (of the 1 to 2 mentioned by Hoover) actually responded, and all of them reported on 1/3 of the items. Half of the respondents are assumed to accurately report the average price over the year, equivalent to 12 monthly observations. The other half reported only one monthly observation. In this scenario, the vast majority of years are reported, which is in line with the fact that the survey was mostly complete for the last 5 to 10 years of the survey.

The number of observations tumbles to just over 2,000 each year under a pessimistic scenario, which reflects the situation for the early period of the Weeks Report. I set the share of items reported at 1/4. This is in line with the idea that most respondents reported on less than 1/3 of all items. Moreover, I take into account that many of the annual averages reported were informed guesses based on partially available information rather than averages of accurate monthly information. I thus reduce the share of implicit monthly observations to 1/4 and assume that the remaining 3/4 reported one representative monthly price quote. Finally, I assume that only half of all annual observations were reported, which is line with the fact that the data are particularly scanty at the beginning of the sample period.

The number of underlying price quote observations was significantly lower for retrospective estimates of a 19th CPI. Moreover, the number of observations substantially increased starting in the early 20th century. This suggests that sampling error, and the associated errors-in-variables problem, is especially relevant when analyzing historical CPI data.

Appendix C. Robustness and specification tests

Table C.1. First-stage specification tests

	T.3. (A)	T.3. (B)	T.4. (A)	T.4. (B)	T.5. (A)	T.5. (B)	T.5. (C)
rk *LM*-statistic	18.52	16.12	28.13	19.06	5.63	4.73	12.15
p-value	0.00	0.00	0.00	0.00	0.02	0.03	0.00
rk *F*-statistic	31.88	25.26	44.51	13.34	7.18	34.41	7.83
5% critical value for 10% of worst-case-bias	23.10	23.10	23.10	7.03[a]	23.10	23.10	7.03[a]
Observations	91	88	143	143	88	143	143

Note: The rk *LM*-statistic tests whether the model is identified in the presence of heteroscedasticity and autocorrelation (Kleibergen and Paap, 2006). The rk *F*-statistic tests whether the model is only weakly identified. Critical values calculated according to Montiel Olea and Pflueger (2013). The worst-case-benchmark corresponds to OLS when errors are conditionally homoscedastic and serially uncorrelated. Note that Montiel Olea and Pflueger (2013) do not report critical values for more than one endogenous regressor. Therefore, critical values with superscript (a) are those from Stock and Yogo (2005) for the case of i.i.d errors.

Table C.2. Brackets on δ

	(1) ΔGDP_t	(2) ΔIP_t	(3) $\widehat{GDP}_t$	(4) $\widehat{IP}_t$
	(A) CPI inflation (1800-1945)			
Lower	[-4.47, -1.12]	[-6.44, -0.09]	[-5.55, -1.84]	[-9.15, -2.83]
Upper	[-48.42, -18.90]	[-200.75, -12.01]	[-45.49, -19.41]	[-76.43, -33.91]
R-squared	0.08	0.03	0.11	0.11
Observations	146	146	146	146
	(B) Wholesale price inflation (1800-1945)			
Lower	[-5.22, -2.22]	[-10.40, -4.73]	[-5.24, -1.53]	[-10.03, -4.09]
Upper	[-30.21, -17.54]	[-61.64, -29.26]	[-57.77, -14.03]	[-68.08, -26.94]
R-squared	0.23	0.20	0.12	0.17
Observations	143	143	143	143

Note: The lower bracket is the 95% confidence interval of the OLS estimator of δ in the regression $y_t = c + \delta 1\{\tilde{\pi}_t < 0\} + \nu_t$. The upper bracket is the 95% confidence interval of $1/\gamma$ from the reverse regression $1\{\tilde{\pi}_t < 0\} = c + \gamma y_t + \nu_t$ calculated using the Delta method. The indicator function $1\{\tilde{\pi}_t < 0\}$ assumes unity if CPI inflation is negative and zero otherwise. Panel (B) uses the wholesale price inflation rate instead of the CPI inflation rate. Columns (1)-(2) give the results for GDP and industrial production in growth rates, and columns (3)-(4) give the results for the HP-filtered measures. Confidence bounds are based on HAC-robust standard errors.

Table C.3. Robustness and specification tests US data (1800-1890)

	(1) ΔGDP_t	(2) ΔGDP_t	(3) ΔIP_t	(4) ΔIP_t	(5) $\widehat{GDP}_t$	(6) $\widehat{GDP}_t$	(7) $\widehat{IP}_t$	(8) $\widehat{IP}_t$
(A) Higher deflation threshold								
$1\{\tilde{\pi}_t < 1\}$	-0.65	-5.57**	0.06	-11.47*	-1.98*	-8.40***	-0.65	-5.57**
	(0.57)	(2.05)	(1.65)	(4.44)	(0.81)	(2.35)	(0.57)	(2.05)
Estimator	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
rk *LM*-statistic		14.02		14.02		14.02		14.02
p-value		0.00		0.00		0.00		0.00
rk *F*-statistic		20.52		20.52		20.52		20.52
Observations	88	88	88	88	88	88	88	88
(B) Inflations and deflations in [-5%, 5%]								
$1\{\tilde{\pi}_t < 0\}$	-1.32	-4.61*	-2.09	-9.36*	-2.04*	-6.55**	-2.79	-10.78*
	(0.78)	(1.99)	(1.59)	(4.07)	(0.90)	(2.25)	(1.73)	(4.06)
Estimator	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
rk *LM*-statistic		9.78		9.78		9.78		9.78
p-value		0.00		0.00		0.00		0.00
rk *F*-statistic		13.22		13.22		13.22		13.22
Observations	61	61	61	61	61	61	61	61
(C) CPI deflations and common deflations signaled by CPI as well as the proxy								
$1\{\tilde{\pi}_t < 0\}$	-1.53*	-2.14**	-2.00	-4.08**	-2.42**	-3.73***	-3.01*	-5.10***
	(0.66)	(0.67)	(1.57)	(1.38)	(0.76)	(0.73)	(1.37)	(1.39)
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Deflations	CPI	Common	CPI	Common	CPI	Common	CPI	Common
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	88	60	88	60	88	60	88	60
(D) Identification in reverse regression (Hahn and Hausman, 2002)								
$1\{\tilde{\pi}_t < 0\}$	-1.26	-3.60*	-1.69	-8.19**	-2.20**	-6.16***	-2.63	-9.03***
	(0.65)	(1.45)	(1.45)	(3.00)	(0.72)	(1.54)	(1.35)	(2.61)
Estimator	OLS	RIV	OLS	RIV	OLS	RIV	OLS	RIV
Controls	No	No	No	No	No	No	No	No
Observations	91	91	91	91	91	91	91	91
(E) Use CPI inflation as instrument for proxy								
$1\{z_t < 0\}$	-2.24***	-3.27*	-4.84***	-4.28	-3.62***	-5.18***	-5.34***	-6.45*
	(0.65)	(1.30)	(1.34)	(3.04)	(0.73)	(1.35)	(1.23)	(2.58)
Estimator	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
rk *LM*-statistic		16.12		16.12		16.12		16.12
p-value		0.00		0.00		0.00		0.00
rk *F*-statistic		23.69		23.69		23.69		23.69
Observations	88	88	88	88	88	88	88	88

Note: For Panels (A)-(B), see Table 3. Panel (C) performs OLS estimates limiting the samples to deflations and inflations signaled by both, the error-ridden CPI and the proxy variable. Following Hahn and Hausman (2002), Panel (D) reports OLS estimates of $\delta = 1/\gamma$ of the reverse regression of the deflation dummy on real activity: $1\{\tilde{\pi}_t < 0\} = \alpha + \gamma y_t + \epsilon_t$. In the reverse IV-regressions (RIV) real activity is instrumented using the deflation dummy based on the proxy. Panel (E) uses the proxy (z_t) to classify deflations and uses the CPI inflation rate as an instrument.

Table C.4. Placebo and specification tests US data (1957-2015)

	(1) ΔGDP_t	(2) ΔGDP_t	(3) ΔIP_t	(4) ΔIP_t	(5) $\widehat{GDP}_t$	(6) $\widehat{GDP}_t$	(7) $\widehat{IP}_t$	(8) $\widehat{IP}_t$
(A) Exclusion restriction wholesale price dummy								
$1\{\pi_t < 3\}$	0.19	0.24	1.08	1.34	-0.75	-0.11	-0.93	-0.14
	(0.60)	(0.77)	(1.25)	(1.26)	(0.68)	(0.79)	(1.21)	(1.27)
$1\{\Delta wpi_t < 3\}$		-0.09		-0.51		-1.29		-1.58
		(0.70)		(1.13)		(0.75)		(1.17)
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	59	59	59	59	59	59	59	59
(B) Exclusion restriction wholesale price inflation								
π_t	-0.16	-0.36*	-0.36	-1.02**	0.08	-0.06	0.21	-0.09
	(0.12)	(0.16)	(0.26)	(0.32)	(0.13)	(0.20)	(0.24)	(0.34)
Δwpi_t		0.16		0.51*		0.11		0.24
		(0.10)		(0.22)		(0.11)		(0.20)
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	59	59	59	59	59	59	59	59

Note: The table gives placebo tests and tests for the exclusion restriction based on modern US data from 1957-2015. The regressions control for share price inflation.

Table C.5. Real economic performance during deflations in more countries

	(1) ΔGDP_{it}	(2) ΔGDP_{it}	(3) $\widehat{GDP}_{it}$	(4) $\widehat{GDP}_{it}$
(A) 17 countries (1870-1945)				
$1\{\tilde{\pi}_{it} < 0\}$	-0.15	-3.43	-2.68***	-11.24**
	(0.48)	(2.66)	(0.72)	(3.49)
Estimator	OLS	IV	OLS	IV
rk *LM*-statistic		18.68		18.68
p-value		0.00		0.00
rk *F*-statistic		21.47		21.47
Countries	17	17	17	17
Observations	703	703	703	703
(B) Excluding US (1870-1945)				
$1\{\tilde{\pi}_{it} < 0\}$	0.17	-3.08	-2.33**	-9.79**
	(0.47)	(2.55)	(0.72)	(3.27)
Estimator	OLS	IV	OLS	IV
rk *LM*-statistic		18.30		18.30
p-value		0.00		0.00
rk *F*-statistic		21.17		21.17
Countries	16	16	16	16
Observations	648	648	648	648
(C) With interaction term (1870-1945)				
$1\{\tilde{\pi}_{it} < 0\}$	-0.62	-1.11	-1.55**	-8.41*
	(0.50)	(3.77)	(0.56)	(3.60)
$1\{\tilde{\pi}_{it} < 0\} \times$	-0.72	-1.20	-1.83	1.13
$1\{t > 1914\}$	(0.88)	(4.13)	(1.17)	(4.16)
Estimator	OLS	IV	OLS	IV
rk *LM*-statistic		6.63		6.63
p-value		0.01		0.01
rk *F*-statistic		4.50		4.50
Countries	17	17	17	17
Observations	703	703	703	703

Note: See also Table 3. The list of countries is given in Appendix A. The IV-regressions use a dummy based on lagged broad money growth as an instrument for the deflation dummy. All regressions include country-time fixed-effects. Other control variables include real house price as well as equity price inflation, and a systemic crisis dummy. Coefficients on constant and controls are not shown. The dependent variables are real per capita GDP growth and an HP-filtered output gap. HAC-robust standard errors are given in parentheses. Coefficients with superscripts ***/**/* are statistically significant at the 1%/5%/10% level. Annual data for 17 countries stem from Jordà et al. (2016), and Knoll et al. (2016).

Figure C.1. Euro area inflation

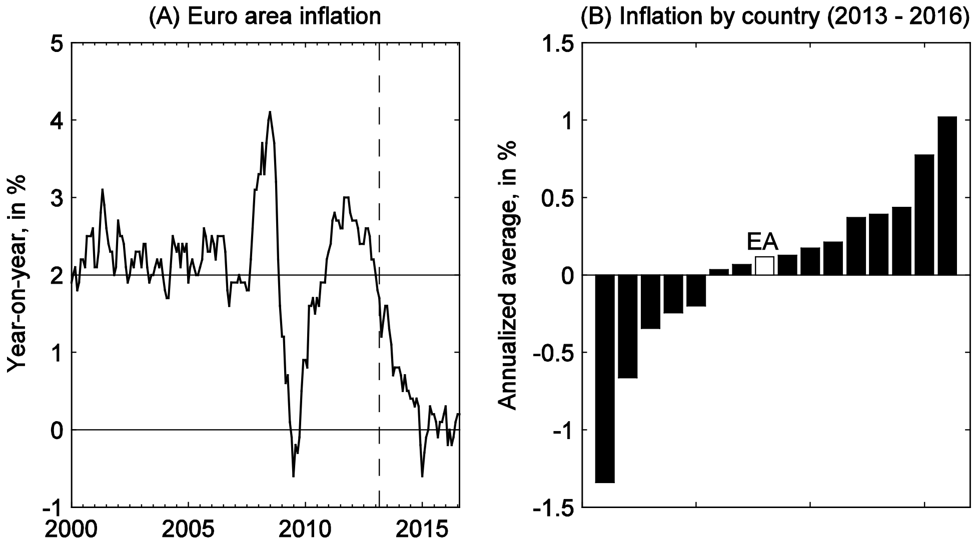


Table C.6. Real economic performance during deflations in the Euro Area

	(1) ΔGDP_{it}	(2) ΔGDP_{it}	(3) $\widehat{GDP}_{it}$	(4) $\widehat{GDP}_{it}$	(5) $Urate_{it}$	(6) $Urate_{it}$	(7) $\widehat{Urate}_{it}$	(8) $\widehat{Urate}_{it}$
				(A) 15 Euro Area countries				
$1\{\pi_{it} < 0\}$	0.63	0.23	-2.47*	0.00	2.87*	0.60	2.30**	0.42
	(0.53)	(0.68)	(1.13)	(0.83)	(1.13)	(1.03)	(0.79)	(0.67)
π_{it}		-0.01		1.50***		-0.98***		-0.89***
		(0.16)		(0.29)		(0.30)		(0.21)
$\pi_{it} \times 1\{\pi_{it} < 0\}$		-0.58		0.91		-1.63		-1.20
		(0.91)		(1.30)		(1.50)		(0.96)
R-squared	0.82	0.82	0.69	0.77	0.50	0.58	0.54	0.64
Observations	133	133	125	125	133	133	133	133
				(B) Higher threshold				
$1\{\pi_{it} < 1\}$	0.26	0.32	-1.90	0.01	2.20*	1.29	1.54	0.61
	(0.71)	(0.83)	(1.31)	(1.29)	(1.00)	(0.99)	(0.82)	(0.76)
π_{it}		-0.02		1.52***		-0.91**		-0.86***
		(0.17)		(0.29)		(0.30)		(0.22)
$\pi_{it} \times 1\{\pi_{it} < 1\}$		-0.38		0.64		-1.47		-1.11*
		(0.46)		(0.80)		(0.83)		(0.55)
R-squared	0.81	0.82	0.68	0.77	0.47	0.58	0.50	0.64
Observations	133	133	125	125	133	133	133	133

Note: The table shows regressions of the form $y_{it} = c_i + d_t + \beta_1\pi_{it} + \beta_2\pi_{it} \times 1\{\pi_{it} < 0\} + \delta 1\{\pi_{it} < 0\} + \theta' controls_{it} + \nu_{it}$. The indicator function $1\{\pi_{it} < 0\}$ assumes unity if CPI inflation is negative and zero otherwise. All regressions include country-time fixed-effects. Other control variables include real house price as well as equity price inflation. Coefficients on constant and controls are not shown. The dependent variables are (1)-(2) GDP growth, (3)-(4) an output gap, (5)-(6) the unemployment rate, (7)-(8) the unemployment gap. HAC-robust standard errors are given in parentheses. Coefficients with superscripts ***/**/* are statistically significant at the 1%/5%/10% level.