University of Neuchatel

Institute of Economic Research

IRENE Working paper 17-09

Is Deflation Costly After All? The Perils of Erroneous Historical Classifications

Daniel Kaufmann

UNIVERSITÉ DE NEUCHÂTEL

Institut de recherches économiques

Is Deflation Costly After All? The Perils of Erroneous Historical Classifications *

Daniel Kaufmann†

First version: 13 July 2016
This version: 1 September 2017

Abstract: Measurement error in historical data distorts descriptive analyses based on binary classifications. Modern replications of deficiencies in retrospective CPI estimates for the 19th century show that measurement issues cause misclassification of inflationary and deflationary episodes. We therefore underestimate the shortfall in real activity during deflation. Using various approaches to control for measurement error in 19th century US CPI data, a series of stylized facts emerge: (i) Real activity was on average substantially lower during deflations; (ii) CPI deflations were associated with at least as severe shortfalls in real activity as equity price declines and banking crises; (iii) Only severe deflations were associated with declines in real activity; (iv) Transitory and persistent deflations, as well as, monetary and nonmonetary deflations were equally associated with lower GDP growth.

JEL classification: E31, E32, N11, C2

Keywords: Deflation, real activity, monetary history, measurement error, binary regressors, misclassification bias, bounds, GMM.

*This project strongly benefited from a visit at the Berkeley Economic History Laboratory (BEHL), whose hospitality I gratefully acknowledge. I thank Christiane Baumeister, Gregor Bäurle, Bernd Bartels, Gillian Brunet, Brad DeLong, Barry Eichengreen, Stefan Gerlach, Yuriy Gorodnichenko, Savina Gygli, Philipp Harms, Bo Honoré, Matthias Hölzlein, Florian Huber, Ronald Indergand, Jan Jacobs, Dmitri Koustas, Carlos Lenz, Petros Milionis, Andreas Müller, Klaus Neusser, Nuno Palma, Steve Pischke, Tobias Renkin, Christina Romer, Gisela Rua, Samad Sarferaz, Jan-Egbert Sturm, Michael Siegenthaler, Eric Sims, Zach Stangebye, Richard Sutch, John Tang, Cédric Tille, Michael Weber, Jonathan Wright, and Kaspar Wüthrich, as well as seminar participants at UC Berkeley, the University of Notre Dame, the Federal Reserve Board, the University of Mainz, the EEA-ESEM, the Swiss National Bank, the University of Groningen, WU Vienna, the KOF/ETH Zurich, and the SSES Annual Meeting, 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 US price 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

Historical studies often report descriptive statistics conditional on categorical classifications based on retrospectively estimated macroeconomic data. To name a few recent examples, Reinhart and Rogoff (2010) estimate average GDP growth conditional on various ranges of public debt as a share of GDP, Jorda et al. (2016) estimate average real activity during periods with high and low credit growth, and Borio et al. (2015) examine average GDP growth during inflation and deflation. The latter two can be cast into a framework of regressing real activity on a binary indicator variable. If this binary indicator is measured with error, however, OLS will suffer from a misclassification bias (see Aigner 1973). To see this, let us assume that deflation is actually associated with low GDP growth. If we use a mismeasured price index to classify deflations, some of those periods will actually be associated with rising prices and therefore high GDP growth. The average calculated conditional on the error-ridden deflation indicator will therefore necessarily suffer from an upward bias.

We know from studies that replicate the methodological deficiencies of historical macroeconomic series based on modern data sources that those deficiencies distort the time-series properties of real activity measures (see Romer 1986a,b), as well as, nominal wage, wholesale price and CPI inflation (see Allen 1992, Hanes 1998, Kaufmann 2017). Providers of historical data therefore sometimes warn that such data are not suitable for sophisticated time-series analysis (see Johnston and Williamson 2008, for historical GDP data). Little is known, however, whether measurement issues hamper even seemingly simple descriptive analyses, such as measuring the average state of the real economy during deflation.

This paper sheds light on the misclassification bias arising when measuring the association between real activity and deflation (see e.g. Borio et al. 2015, Eichengreen et al. 2016). It provides estimates on the average shortfall in real activity during 19th century deflations in the US using a bias-adjustment, as well as, three proxy variable approaches that can be used to derive bounds, sets, and point estimates. The former use misclassification rates derived from comparing well-measured post-WWII price indices to replications that mimick the methodologies applied by economic historians to retrospectively construct 19th century CPI data (see Kaufmann 2017). The latter build on Kane et al. (1999) and Black et al. (2000) who deal with misreported categorical response variables typically encountered in modern survey data. By focusing on a reduced-form regression framework, the analysis remains silent on the direction of causation. Examining the impact on structural analysis, for example along the lines of Bayoumi and Eichengreen (1996), Bordo and Redish (2004) and Beckworth (2007), is beyond the scope of this paper.[1]

The main findings can be summarized as follows. After accounting for measurement error, four interesting stylized facts emerge: (i) GDP growth fell on average by 2.4 percentage points during deflations. This decline is about 0.7 percentage point stronger compared to the OLS estimate and similar across various approaches. Using an output gap measure or industrial production growth to measure real activity, the decline is even more pronounced. (ii) CPI deflations were associated with at least as severe shortfalls in GDP growth than equity price declines and banking crises. While equity price declines were associated with 1.9 percentage points lower GDP growth, there was no significant association with banking crises. (iii) Only severe deflations, that is deflations with declines in the price level of more than 3%, were associated with significant declines in real activity; (iv) Transitory and persistent deflations, as well as, monetary and nonmonetary deflations were associated with lower GDP growth and there were no significant differences among those different types of deflations.

This paper is related to a wealth of studies on the real effects of deflation (see e.g. Atkeson and Kehoe 2004, Bordo and Filardo 2005, Borio et al. 2015, Eichengreen et al. 2016). This question has become more relevant recently against the backdrop of nonconventional policy actions by central banks that are at least partially justified by an imminent threat of deflation. The studies use large cross-country panels including data from the 19th century. The main reason to use data from the distant past is that the 19th century comprises many interesting case studies. Regular 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). For

[1] But, characterizing the nature and severity of measurement error is likely informative for dynamic structural analysis. Komunjer and Ng (2014) discuss identifiability of dynamic models when measurement errors are persistent. Moreover, Mumtaz et al. (2015) show that some strategies for identifying credit supply shocks perform poorly in the presence of measurement error.

example, during the 19th century US, the consumer price level declined nearly half of the time and, on average over all deflations, the price level declined by 4.5% during a deflationary episode. This shows that deflation was not only frequent but also substantial.

The controversial findings on the subject are in line with measurement issues in price data. Atkeson and Kehoe (2004), Bordo and Filardo (2005), and Borio et al. (2015) 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. One explanation for the weak link is that many 19th century deflations were benign in the sense that they were driven by advances in productivity and therefore associated with high real income growth (see Friedman and Schwartz 1963, Beckworth 2007). Another reason is that less rigid product and labour markets implied that 19th century economies adapted more rapidly to adverse deflationary demand shocks and thus the real consequences were less severe (see Bayoumi and Eichengreen 1996, and references therein). This paper argues in favour of a third possibility, namely that measurement issues in historical price data attenuates the empirical link between real activity and deflation and misclassification error is particularly severe for 19th century CPI inflation.

The remainder of the paper is structured as follows. I first elaborate on the misclassification bias arising from measurement error in binary variables and discuss various approaches to recover the actual association between real activity and deflation. Then, I discuss the data sources. The next section presents estimates of the shortfall of real activity during deflationary episodes for the 19th century US. The last section concludes.

2 Econometric background

Against the backdrop of a reduced-form regression framework this section characterizes the misclassification bias resulting from estimating the average of a variable conditional on an erroneous binary indicator. Although dealing with measurement error is somewhat more involved in the binary than in the continuous case, the framework is relevant for at least two reasons. Many descriptive analyses using historical data report averages conditional on a potentially mismeasured classification. Moreover, the main question asked in this paper implies that there is a discontinuity in the relationship between inflation and real activity when inflation falls below zero.

2.1 Characterizing the misclassification bias

Researchers have examined the link between real activity and deflation by regressing GDP growth on a deflation indicator (see e.g. Borio et al. 2015, Eichengreen et al. 2016). In the simplest case of only one country and no additional control variables, the regression equation reads:

$$y_t = \alpha + \beta d_t + \varepsilon_t \ , \tag{1}$$

where y_t is a measure of real activity and $d_t \equiv \mathbf{1}_{\{\pi_t<0\}}$ is an indicator variable assuming unity if inflation is negative and zero otherwise. Moreover, ε_t is an *i.i.d.* error term capturing unexplained factors including classical measurement error in the real activity variable.[2] The reduced-form OLS estimate is then used to examine whether deflation is associated with lower real activity. A negative coefficient on the deflation dummy indicates that real activity has been on average lower during deflationary episodes than during inflationary episodes. This framework is relevant more broadly because studies that report an average according to a binary classification report point estimates of α and $\alpha+\beta$.

Unfortunately, retrospectively constructed measures of inflation are likely measured with error (see Hanes 1998, Kaufmann 2017). We are therefore estimating

$$y_t = \alpha + \beta x_t + \epsilon_t \ , \tag{2}$$

with $x_t \equiv \mathbf{1}_{\{\tilde{\pi}_t<0\}}$ and $\tilde{\pi}_t = \rho_0 + \rho_1\pi_t + \omega_t$. Following Kane et al. (1999), the binary indicator is based on a mismeasured proxy that is linearly related to the true inflation rate and suffers from an *i.i.d.* measurement error ω_t (see Kane et al. 1999). Consequently, the mismeasured dummy x_t will classify some periods as deflations, when prices were actually rising, and some periods as inflations when prices were in fact falling. Following

[2] OLS is still consistent in the presence of classical measurement error in the dependent variable although the estimates will become less precise (see e.g. Hausman 2001).

Aigner (1973) we can assume that $\pi_t, \omega_t, \varepsilon_t$ are mutually independent and show that the OLS estimate of β is biased:[3]

$$\begin{aligned} plim\ \hat{\beta}_{OLS} &= E[y_t|x_t = 1] - E[y_t|x_t = 0] \\ &= \alpha + \beta P[d_t = 1|x_t = 1] - \alpha - \beta P[d_t = 1|x_t = 0] \\ &= \beta(1 - P[d_t = 0|x_t = 1] - P[d_t = 1|x_t = 0]) \ . \end{aligned} \tag{3}$$

Because probabilities are positive, it follows that the OLS estimate is biased upwards (downwards) if β is negative (positive). If the dummy misclassifies a high share of inflations and deflations, the OLS estimate may even be of the wrong sign. The bias is zero if x_t correctly classifies all inflations and deflations.

The bias has an intuitive interpretation. Assume that deflation is actually associated with lower real activity but we use a mismeasured price index to classify deflationary and inflationary episodes. If we calculate average real activity during deflations, some of those periods were in fact associated with rising prices and relatively high real activity. Therefore, average real activity based on the erroneous classification will wrongly include some inflationary episodes and therefore overestimate real activity during deflations. By contrast, if we calculate average real activity during inflations, some of them were actually associated with falling prices and low real activity. Therefore, we will underestimate average growth during inflationary periods.

How does the misclassification bias differ from the continuous case, where we regress real activity on the inflation rate directly? Recall that the probability limit of the regression coefficient in the continuous case amounts to (see e.g. Griliches 1986):

$$plim\ \hat{\beta}_{OLS} = \beta \frac{\rho_1 \sigma_\pi^2}{\rho_1^2 \sigma_\pi^2 + \sigma_\omega^2} \ , \tag{4}$$

Under the classical assumptions $\rho_1 = 1$ and the bias only depends on the signal to noise ratio, that is, the relative volatility of the correctly measured inflation rate and the measurement error process. In this case, because variances are positive, the attenuation

[3] Pakes (1982) shows that Wald-type estimators are generally biased if the underlying data are subject to error. In such estimators, we divide observations in groups with above and below median observations on the independent variable and then fit a line through the group means.

factor lies between zero and unity and therefore the OLS estimate will be attenuated towards zero. If the variance of the measurement error tends to zero, the attenuation factor tends to unity and OLS is unbiased.

Data transformations, for example squaring the variable, sometimes lead to a more severe bias (see Griliches 1986). Forming a binary indicator also alters the bias relative to the continuous case.[4] To gain some insights, I simulate the misclassification and attenuation factors for the binary and continuous case, respectively, where the inflation rate and measurement error are assumed to be *i.i.d.* normally distributed.[5]

Panels (A) and (B) in Figure 1 provide simulations with classical measurement error in the mismeasured inflation rate ($\rho_1 = 1$). In Panel (A) the overall volatility of the error-ridden inflation rate is set to $\sqrt{\sigma_\pi^2 + \sigma_\omega^2} = 6$, which corresponds to the volatility of CPI inflation in the 19th century US, and then simulates the misclassification and attenuation factors for various signal to noise ratios. The solid line, representing the attenuation factor for the continuous case, is higher if the signal to noise ratio is larger than one. Thus, for relatively mild classical measurement error, the bias is more pronounced in the binary than in the continuous case.

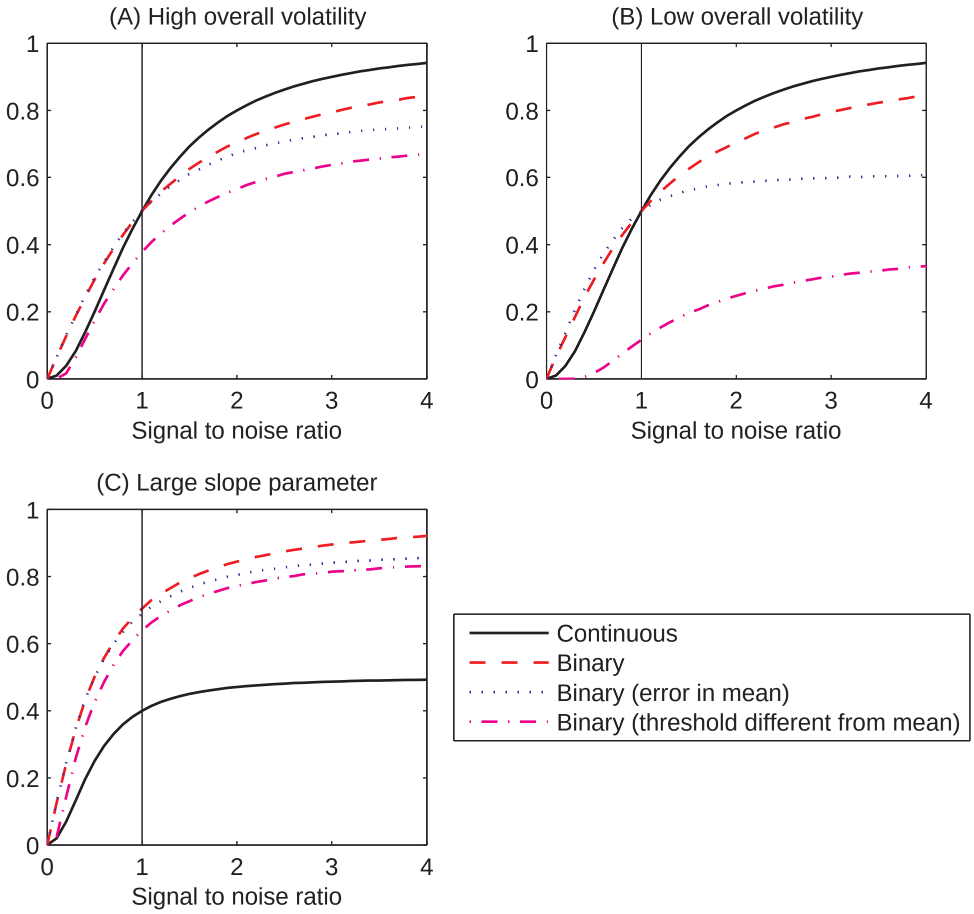
The misclassification bias becomes more severe when we mismeasure the mean of inflation. Setting the mean of the measurement error process to $\rho_0 = 2$ lowers the misclassification factor when the signal to noise ratio is larger than unity (dotted line). This bias does not vanish as the signal to noise ratio approaches infinity because we always misclassify at least some episodes. Interestingly, the bias also depends on the specific threshold value we use to form the binary indicator. If this threshold differs from the mean of inflation, for example 2 instead of 0, the bias becomes more severe (dash-dotted line). Notably, this is even the case for signal to noise ratios smaller than unity. The intuition for this result is that a threshold differing from the mean of the signal renders the event less probable.[6] Observing such a rare event is therefore likely due to a

[4]Kreider (2010) shows that with arbitrary forms of classification error, even moderate rates of misclassfification can lead to very serious biases.

[5]Appendix A derives some analytical results for a more restrictive special case.

[6]This case is particularly relevant for studies reporting averages conditional on categorical variables covering relatively rare events. Reinhart and Rogoff (2010), for example, calculate average GDP growth over long historical episodes and various advanced economies in bins of debt to GDP ratios of below 30%, 30% to 60%, 60% to 90% and above 90%. They report that GDP growth was lower in the highest category than in the other categories. If a debt to GDP ratio of above 90% is a relatively rare event, econometric theory suggests that the misclassification bias is more severe in this particular bin.

Figure 1 — Simulated attenuation and misclassificaiton factors



Note: The figure shows misclassification and attenuation factors for various signal to noise ratios. Their analytic counterparts are given in equations (3) and (4). The mismeasured inflation rate takes the form $\tilde{\pi}_t = \rho_0 + \rho_1 \pi_t + \omega_t$ and the dummy variable is formed as $x_t \equiv \mathbf{1}_{\{\tilde{\pi}_t < c\}}$. The actual inflation rate as well as the additive measurement error (ω_t) are assumed to be normally distributed with zero mean. The overall volatility of the inflation rate and measurement error ($\sqrt{\sigma_\pi^2 + \sigma_\omega^2}$) is fixed at 6 standard deviations (Panels A and C) and 2.6 standard deviations (Panel B). For the binary case, the simulations are shown form measurement error with nonzero mean ($\rho_0 = 2$) and threshold differing from the true mean of inflation ($c = 2 \neq E(\pi_t) = 0$). Panel (C) simulates non-additive measurement error setting the slope parameter to $\rho_1 = 2$ so that the proxy is more volatile than the actual underlying inflation rate.

misclassification.[7]

The bias becomes particularly severe if the overall volatility of the process is low, while, in the continuous case the overall volatility does not affect the size of the bias at all. Panel (B) shows the simulations for an overall volatility of 2.6 standard deviations, which is in line with the post-WWII volatility of the CPI inflation rate. It turns out that if we mismeasure the mean by 2, the misclassification factor drops to 0.6 in the binary case even for high signal to noise ratios. Moreover, if the threshold differs from the mean of the error-free inflation rate the bias becomes even more severe.

Panel (C) shows the results under nonclassical measurement error where I assume that $\rho_1 = 2$ so that the proxy for the inflation rate is necessarily more volatile even absent classical measurement error. The attenuation factor declines in the continuous case. In the binary case, however, the bias becomes less severe as the signal becomes stronger. Therefore, if we obtain a well measured proxy with the only deficiency being that it is too volatile, the classification can be less severe even if the attenuation bias becomes more pronounced.

To summarize, relative to the attenuation bias, the misclassification bias is particularly relevant in the presence of measurement error in the mean, if the threshold does not correspond to the mean of the error-free inflation rate, and if the overall volatility of the process is low. Meanwhile, a proxy whose only deficiency is that it is too volatile can be less affected by the misclassification bias than the attenuation bias.

2.2 Bias adjustments, bounds and point estimates

To resolve the misclassification bias, I make use of four approaches requiring various assumptions. First, as Aigner (1973) suggests, if we obtain information on the misclassification rates we can bias-adjust our estimates using equation (3). The main assumption underlying this approach is that the misclassification rates are accurate.

Second, let us assume that we obtain another error-ridden measurement of the deflation dummy $z_t \equiv \mathbf{1}_{\{\hat{\pi}_t<0\}}$, where $\hat{\pi}_t = \gamma_0+\gamma_1\pi_t+\psi_t$ and $\pi_t, \omega_t, \psi_t, \varepsilon_t$ are mutually independent.[8]

[7] The same intuition applies to medical testing. If the illness to be detected is rare, a small false positive rate may imply that most of the positively tested individuals are healthy.

[8] Typically, this involves two independent surveys on the same binary indicator. For example, Black et al. (2000) obtain independent surveys from employers as well as employees on whether the employee is eligible for health insurance or not.

More specifically, we require that the two deflation indicators are independent conditional on the actual outcome of the true deflation dummy d_t. Then, Black et al. (2000) propose a simple way to estimate a bound that, even in finite samples, is closer to the true parameter than the individual OLS estimates. The OLS estimate of β_{11} in

$$y_t = \alpha + \beta_{11}\mathbf{1}_{\{x_t=1,z_t=1\}} + \beta_{10}\mathbf{1}_{\{x_t=1,z_t=0\}} + \beta_{01}\mathbf{1}_{\{x_t=0,z_t=1\}} + \epsilon_t \qquad (5)$$

will in expectation be closer to the true value of β. In this regression, β_{11} measures the association between real activity and deflation for episodes during which both proxies decline. Intuitively, if we obtain two independent signals that a period was indeed associated with deflation, the probability of a misclassification is lower and therefore the bias smaller. The OLS estimate will still be biased, however, because it is possible that both indicators misclassify a deflation period at the same time. Therefore, this approach yields a conservative estimate of the shortfall in real activity during deflations.

With classical measurement error, we can use the second proxy as an instrument (see Hausman 2001). It is worth noting, however, that the classical assumptions are violated in the binary model because the classification error is necessarily negatively correlated with the outcome (see Kane et al. 1999). If $d_t = 1$, the classification error can only amount to 0 or -1, whereas if $d_t = 0$, the classification error amounts to either 0 or 1. Kane et al. (1999) show that IV will actually deliver a biased estimate in the opposite direction:

$$plim\ \hat{\beta}_{IV} = \frac{\beta}{1 - P[x_t = 0|d_t = 1] - P[x_t = 1|d_t = 0]} \ .$$

The IV estimate therefore provides a lower bound if $\beta < 0$.[9]

Third, under the same assumption, we can consistently estimate the coefficient using GMM (see Kane et al. 1999, Black et al. 2000). We can calculate three independent sampling fractions conditional on each combination of outcomes of the two binary indicators, as well as, four averages of the real activity variable conditional on each combination of outcomes. From these empirical moments, we have to estimate seven parameters (two model coefficients, four misclassification rates, and the true rate of

[9]Note that the IV estimates reported in the working paper version of this paper therefore represent a lower bound (see Kaufmann 2016).

deflation) such that the model is just identified. Appendix A provides more details and shows that the model can be extended to interaction terms and additional well-measured binary covariates. In the latter case, it turns out the model is over-identified.

Without the conditional independence assumption the parameter of interest is not point identified. From the seven empirical moments we can estimate from the data, we have to identify nine parameters (see Appendix A). As a fourth approach, this paper uses extraneous information derived from the modern replications and explores estimates of β fixing two of the unknown parameters at reasonable values.[10] To be concrete, I fix the probabilities that the two indicators jointly misclassify a deflationary ($P[x_t = 0, z_t = 0|d_t = 1]$) or an inflationary episode ($P[x_t = 1, z_t = 1|d_t = 0]$). If measurement error in the two indicators is not too severe, those probabilities are small.[11] If we are willing to assume that the two joint misclassification probabilities are smaller than a certain value, our parameter of interest is set identified.

The four approaches have several advantages and disadvantages. Bias-adjustment hinges on the assumption that we can obtain accurate information on the misclassification rates but does not require an additional error-ridden indicator. The bounding approach and GMM have the advantage that we do not need to know the misclassification rates. In fact, GMM will deliver consistent estimates thereof (see Appendix A). We have to assume, however, that the error-ridden indicators are conditionally independent of each other and that the measurement error is uncorrelated with the actual inflation rate and other covariates. The bounding approach is particularly useful compared to GMM in finite samples. In fact, at least in the present application, GMM becomes infeasible if we include multiple covariates as many cells for which we have to calculate sampling fractions contain no observations. By contrast, the bounding approach delivers a conservative estimate even in finite samples. Finally, to relax the conditional independence assumption, we have to obtain additional information to determine sensible ranges to fix the joint misclassification

[10] I would like to thank Bo Honoré bringing this possibility to my attention.

[11] The intuition for this assumption is similar to medical testing. Testing whether a patient has a rare illness can be subject to testing error. However, if two separate tests diagnose the illness, the probability of jointly misdiagnosing the illness is small and we have greater confidence in the result. The assumption underlying the identification strategy is that the joint misclassification error is relatively small while the individual misclassification error can still be high.

probabilities.[12] But even then, the coefficient of interest is only set identified rather than point identified. Because of those advantages and disadvantages, the remainder of the paper will provide estimates based on all three approaches.

3 Data

The analysis uses historical US data from 1800-1899. To form a binary deflation indicator, I use the composite annual CPI by Officer and Williamson (2016). This index is based on a careful selection of alternative retrospective estimates of CPI inflation as discussed by Officer (2014). Although this series likely represents the most accurate estimate at any given point in time, it suffers from various methodological deficiencies which can be traced back to scarce retail data (see Kaufmann 2017). Those measurement problems lead to misclassification of inflationary and deflationary episodes and therefore give rise to a misclassification bias.

To take into account potential non-classical measurement error in the dependent variable, I consider various measures of real activity: Real per capita GDP growth, industrial production growth, both in percent, and percentage deviations of the two variables from their trends. 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 modern industrial production series from the Board of Governors of the Federal Reserve System starts only in 1919. I bridge this gap using the manufacturing production series by Fabricant (1940). The gap measures are calculated using the procedure by Hamilton (2016).[13] For detrending the real activity series, I use a sample spanning from 1790-2015 to avoid potential end-of-sample instabilities.

[12]An additional advantage of the bounding approach is that it scales to more than two indicators. Adding a third indicator, the joint probability of misclassification can not increase relative to two indicators. By contrast, for the conditional independence approach, we have to establish independence of the measurement error for each of the three indicators.

[13]Hamilton (2016) proposes to use the residual of the regression $y_t = \beta_0 + \beta_1 y_{t-s} + \beta_2 y_{t-s-1} + \beta_3 y_{t-s-2} + \beta_4 y_{t-s-3} + \varepsilon_t$ as a gap measure and recommends to set $s = 2$ for annual data. The results are robust to using a Hodrick-Prescott-filter, following Davis et al. (2009), using a smoothing parameter set to 100.

3.1 Misclassification rates from modern replications

Misclassification rates to bias-adjust the OLS estimates are based on replications of 19th century CPIs for the post-WWII US. The data stem from Kaufmann (2017), where I construct CPIs for the period 1956-2016 that are affected by the same methodological deficiencies as the 19th century segments of the composite CPI by Officer and Williamson (2016). Because we observe both, the erroneous as well as the well-measured inflation series, we can gauge the misclassification bias and use this information to bias-adjust the OLS estimates.[14]

The measurement issues vary over the various segments. For simplicity, the misclassification rates are calculated using a replication of a typical retrospective estimate of the period 1860-1880. In this particular segment, rents are approximated by a reproduction cost index, that is an average of wholesale prices for building materials and wages of low-skilled workers, services prices are lacking, and the number of individual price quote observations to compute the CPI is small. Those deficiencies are replicated by using the modern BLS CPI less services, replacing rent with a replication of a reproduction cost index, and adding classical measurement error.[15]

Panel (A) of Table 1 provides descriptive statistics for official CPI inflation series and the modern replication of a typical 19th century CPI. The statistics include the mean (μ), standard deviation (σ) and the persistence of inflation ($\rho(1)$).[16] Compared to the official CPI inflation series, the replication exhibits a lower mean, higher volatility and lower persistence. The latter is in line with the fact that classical measurement error attenuates measures of persistence, for example, the autoregressive parameter in the AR(1) model (see Staudenmayer and Buonaccorsi 2005).

The remaining columns of the table show to what extent the different time series properties give rise to attenuation and misclassification biases. I calculate two measures

[14] That the modern CPI is measured without error is not exactly right. First, the Boskin Commission (1996) shows that the CPI may underestimate actual CPI inflation because of neglected changes in quality. However, the sampling error in modern CPIs is small (see Shoemaker 2014).

[15] The last deficiency is simulated by 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 (see Kaufmann 2017).

[16] 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.

TABLE 1 — MISCLASSIFICATION RATES FOR MODERN REPLICATIONS

(A) CPI inflation (1957-2016)								
	μ	σ	$\rho(1)$	Γ_{c1}	Γ_{c2}	mc^+	mc^-	Γ_b
(1) Official data	3.7	2.8	0.85					
(2) Replication	3.4	3.0	0.73	0.88	0.86	0.00	0.77	0.23
(3) Demeaned replication	-0.4	3.0	0.73	0.88	0.86	0.21	0.10	0.69

(B) PPI inflation (1957-1990)								
	μ	σ	$\rho(1)$	Γ_{c1}	Γ_{c2}	mc^+	mc^-	Γ_b
(1) Official data	4.1	4.8	0.76					
(2) Replication	4.0	5.9	0.47	0.67	0.61	0.00	0.43	0.57
(3) Demeaned replication	-0.2	5.9	0.47	0.67	0.61	0.25	0.10	0.66

Note: Descriptive statistics for official inflation measures and replications of methodological deficiencies in 19th century CPI and WPI data. The modern replications stem from Kaufmann (2017) for the CPI and Hanes (1998) for the WPI. Descriptive statistics include the mean (μ), standard deviation (σ) and the persistence of inflation ($\rho(1)$). The latter 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. For the replications, the table reports two measures of the attenuation factor in the continuous case (Γ_{c1}, Γ_{c2}). The first is the variance of the official inflation measure divided by the variance of the replication. The second is the persistence of the replication divided by the persistence of the official measure. Finally, the table reports the share periods that the indicator wrongly signals an inflation (mc^+) and deflation (mc^-), as well as, the misclassification factor in the binary case ($\Gamma_b = 1 - mc^+ - mc^-$). These are sample analogues of the attenuation factor in equation (3).

of the attenuation factor in the continuous case (Γ_{c1}, Γ_{c2}). The first measure is calculated as the variance of the official inflation rate divided by the variance of the replication. This measure corresponds to the attenuation factor for classical measurement error in equation (4) when $\rho_1 = 1$. The second measure is calculated as the persistence of the replication divided by the persistence of the official inflation series. Finally, the table reports the share of wrongly reported inflations (mc^+) and deflations (mc^-), as well as the misclassification factor in the binary case ($\Gamma_b = 1 - mc^+ - mc^-$). Those measures are sample analogues of equation (3).

The changing time-series properties imply that the CPI wrongly classifies many periods as deflations and thus the misclassification factor falls to 0.23. This implies that we would underestimate the negative association between real activity and deflation by a factor of four. However, this is related to the fact that deflations were an anomaly in the post-WWII era. Recall from the previous section that the bias becomes more severe if the threshold to form the binary indicator differs from the mean of the actual inflation rate. Because average inflation was closer to zero in the 19th century than today, and the threshold value for the deflation indicator is zero, the corresponding misclassification bias for the 19th century is smaller.

Therefore, a conservative measure of the misclassification factor can be calculated by subtracting the well-measured average inflation rate. The misclassification bias is indeed less severe. Nevertheless, the misclassification factor in the binary case (0.69) is still smaller than the attenuation factor in the continuous case (0.88). Because the average inflation rate is now zero, the misclassification bias stems from both, misclassified inflations as well as deflations.

To check the robustness of the result, Panel (B) reports the statistics for the the modern replication of a 19th century wholesale price index by Hanes (1998). The replication is compared to official PPI inflation from the BLS. Similar to the CPI, we observe a slightly lower mean, higher volatility and lower persistence of inflation. But, the attenuation bias is somewhat more severe than in the case of the CPI. By contrast, the misclassification factor is similar as for the CPI.

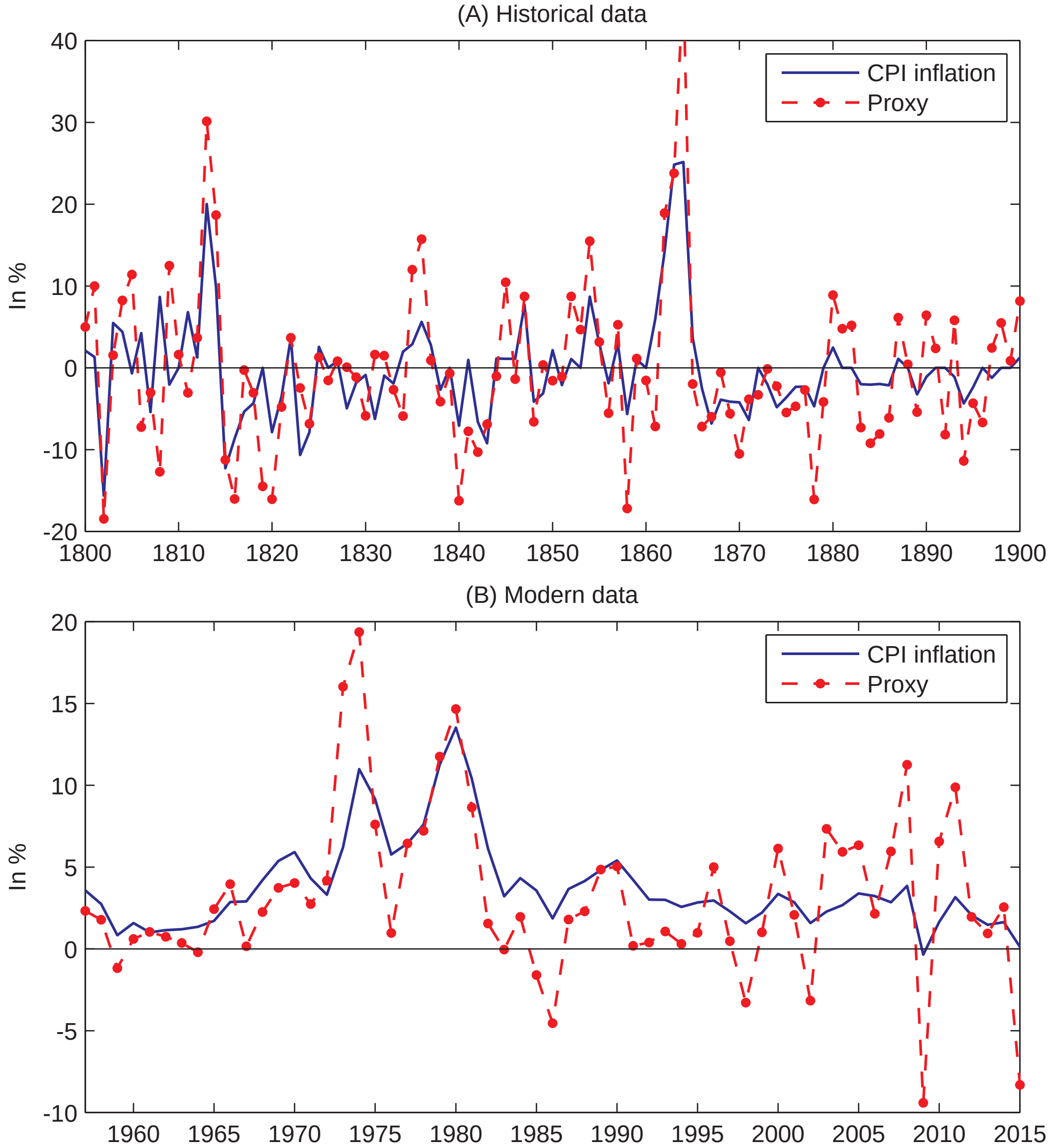
3.2 A proxy variable

The proxy variable approach requires a second, independent, error-ridden measure of CPI inflation. I construct such a proxy variable based on wholesale prices (see Figure 2). For the 19th century, wholesale prices stem from Warren and Pearson (1933) and Hanes (1998). I obtain wholesale prices for the commodity groups food, textile products, fuel and lighting, as well as house furnishings, and aggregate them to a Laspeyres-type index assuming constant consumer expenditure weights by Gordon (2016).[17] The proxy covers approximately 70% of a 19th century consumption basket. 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 impute the series with the weighted average of the other available inflation rates.

Panel (A) of Figure 2 shows the composite CPI inflation rate by Officer and Williamson (2016) as well as the proxy based on wholesale prices from 1800-1900. The two series

[17] I match the Warren and Pearson (1933) commodity groups with the weights from Gordon (2016) 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. See Appendix B for data sources.

Figure 2 — CPI inflation and a proxy



Note: Proxy calculated based on wholesale and producer prices using consumer expenditure weights by Gordon (2016). Composite CPI inflation from Officer and Williamson (2016).

display reasonably similar turning points. Because the proxy is constructed using wholesale prices, it is more volatile. The correlation between the two series, however, is substantial. Interestingly, despite their high correlation, the two variables give different signals concerning deflationary or inflationary episodes. In 25% of all years, the two indicators do not agree on whether it was a deflationary or inflationary episode. This share is surprisingly stable and varies only from 20% to 30% across various episodes.

For the proxy variable approach to be valid, the measurement errors of the proxy based on wholesale prices should be independent from the measurement error of 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 (2016). The Warren and Pearson (1933) data up to 1890 stem from New York newspapers supplemented by prices published in the U.S. Finance Report for 1863 (see Hanes 2006).[18] After 1890, Hanes (1998) provides WPI data consistent with the Warren and Pearson (1933) series based on official BLS data.

The composite CPI is largely based on distinct data sources than the proxy (see Officer 2014, and references therein). 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, Hoover (1960) mainly uses retail prices from the so-called Weeks Report.[19] Partly, 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.[20] The Lebergott (1964) segment from 1860-1880 mainly uses the Weeks Report as well. 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 data because it refers to the difficult period after the Weeks Report. There is no indication that wholesale prices were used. From 1890-1914, the underlying data sources of the composite CPI and the proxy show some overlap.[21] The CPI segment by Rees (1961) from 1890-1914 uses wholesale prices

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

[19] In *1880 Census of the United States, Vol. xx, Joseph D. Weeks, Report on the Statistics of Wages in Manufacturing Industries, with Supplementary Reports.*

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

[21] This concern is addressed by excluding this particular period in a robustness test.

for eleven items from the BLS (1923). Officer (2014) suggests that the index otherwise comprises mainly retail prices. Therefore, the overlap should be modest.

In addition, we can perform some specifications tests using modern data. For the post-WWII era, I construct the proxy using modern PPI data from the BLS.[22] This proxy covers only 13.7% of the consumption basket in 2013. Panel (B) of Figure 2 shows that the proxy is also correlated with post-WWII CPI inflation and reflects major up- and downturns.

Panel (A) of Table 2 tests whether the error-ridden replication as well as the proxy variable are linearly related to the official CPI inflation rate. For the CPI replication and the proxy, the R^2 is higher than 0.5. However, the actual CPI inflation rate is a better proxy because the root-mean-squared error (RMSE) is lower than the RMSE of the proxy. Everything else equal, this would indicate that the bias associated with the CPI inflation rate is less severe than the bias implied by the proxy. In addition, we cannot reject the null hypothesis at the 5% level that the slope coefficient on CPI inflation is unity. For the proxy, the slope coefficient is larger than unity, although, only significant at the 10% level. Note that this is a desirable property because the econometric discussion shows that a slope parameter larger than unity reduces the misclassification bias.[23] For the CPI, the residuals of the regression exhibit significant first-order autocorrelation at the 5% level. For the proxy, we cannot reject the null hypothesis of no serial correlation, at least at the 5% level. These results suggest that the two proxies are sensible, albeit, not ideal indicators.

Serial correlation in the errors is less worrisome than potential correlation of the indicators among each other and with other covariates. The measurement error in the underlying series should be independent of each other, the well-measured inflation rate, as well as of other covariates. Panel (B) shows pairwise rank correlations between CPI inflation, equity prices inflation, which is one of the covariates used in the following application, and the difference between the well-measured CPI and error-ridden inflation rates. The first column shows that the well-measured CPI inflation rate is not significantly

[22]See Appendix B for data sources. The BLS PPI commodity groups are matched with the 2013 weights from Gordon (2016) 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.

[23]In the continuous case it would amplify the attenuation bias.

Table 2 — Specification tests modern data

(A) Linearity assumption

	Replication CPI	Proxy CPI
CPI inflation	0.96***	1.33***
	(0.07)	(0.19)
Constant	-0.29	-1.84**
	(0.23)	(0.75)
RMSE	0.86	3.46
AC(1) (p-value)	0.02	0.06
$\rho_1 = 1$ (p-value)	0.58	0.08
R^2	0.90	0.53
N	59	59

(B) Pairwise rank correlations

	CPI	Equity	Error CPI	Error Proxy
CPI	1.00			
Equity	-0.25*	1.00		
Error CPI	-0.01	0.20	1.00	
Error Proxy	-0.14	0.23*	0.52***	1.00

Note: Panel (A) gives regressions of the CPI replication and the CPI proxy on the well-measured modern CPI inflation rate of the form $\tilde{\pi}_t = \rho_0 + \rho_1 \pi_t + \omega_t$. HAC-robust standard errors are given in parentheses. Coefficients with superscripts ***,**,* are statistically significant at the 1%, 5%, 10% level. AC(1) tests whether the residuals exhibit first-order autocorrelation (see Baum and Schaffer 2013). Panel (B) shows rank correlation coefficients between actual CPI inflation, equity price inflation, and the measurement errors associated with the CPI replication and the CPI proxy. The latter two are calculated as simple differences between the actual CPI inflation rate and the mismeasured inflation rates.

related to the measurement errors of the three inflation measures. The second column, however, shows that the errors are partly correlated with equity price inflation. The correlation is not significant for the CPI replication and only significant at the 10% level for the proxy.

The main issue of the two proxies is that their measurement errors are positively correlated. This is not surprising since both inflation series lack services prices and accurate measures of rent inflation. The positive correlation implies that we only partially recover the actual association and that the estimates based on the bounding strategy and GMM are still conservative. In addition, it emphasizes that we should aim to relax the conditional independence assumption.

4 Estimates for the 19th century

The first set of results show estimates of the shortfall in real activity during deflations in the 19th century US using various approaches assuming conditional independence of the two indicators. Panel (A) of Table 3 shows that 19th century deflations in terms of the CPI were associated with 1.5 percentage points lower GDP growth. The improved upper bound for the association using the additional information from the proxy variable shows a stronger association. GDP growth was on average at least 2.2 percentage points lower during a 19th century deflation. The two point estimates corroborate this finding. The first point estimate, using the misclassification rates to bias-adjust OLS, is almost identical to the improved OLS estimate. Meanwhile, the point estimate using GMM suggests that the shortfall in real activity even amounts to 2.4 percentage points. Finally, the IV estimates suggest that real activity declined at most by 3.7 percentage points.

Using industrial production growth or the two measures of the output gap corroborates this finding. The decline in real activity during a deflationary episode becomes more pronounced according to the improved OLS estimate and point estimate based on GMM. Focusing on the point estimate, industrial production growth declined by 4.2 percentage points and the two output gap measures declined by 5.8 percentage points (GDP) and 9.4 percentage points (industrial production), respectively.

GMM delivers point estimates for the bias implied by the CPI inflation rate and the

Table 3 — Real activity during deflations

(A) GDP growth	Upper bound		Point estimate		Lower bound
Deflation	-1.45**	-2.21***	-2.10**	-2.37***	-3.74**
	(0.72)	(0.77)	(1.04)	(0.81)	(1.54)
Method	OLS	Improved	Adjusted	GMM	IV
Observations	100	100	100	100	100

(B) GDP output gap	Upper bound		Point estimate		Lower bound
Deflation	-3.56***	-5.31***	-5.15***	-5.80***	-8.71***
	(1.05)	(1.17)	(1.52)	(1.37)	(2.40)
Method	OLS	Improved	Adjusted	GMM	IV
Observations	100	100	100	100	100

(C) Industrial production growth	Upper bound		Point estimate		Lower bound
Deflation	-2.06	-4.18***	-2.99	-4.22***	-8.59***
	(1.49)	(1.35)	(2.16)	(1.55)	(3.11)
Method	OLS	Improved	Adjusted	GMM	IV
Observations	100	100	100	100	100

(D) Industrial production output gap	Upper bound		Point estimate		Lower bound
Deflation	-5.92***	-8.78***	-8.58***	-9.36***	-14.63***
	(1.98)	(2.02)	(2.87)	(2.25)	(4.17)
Method	OLS	Improved	Adjusted	GMM	IV
Observations	100	100	100	100	100

Note: Estimates of the model $y_t = \alpha + \beta x_t + \epsilon_t$. The upper bounds of the association are OLS estimates, as well as, the improved OLS estimates by Black et al. (2000) using the proxy as a second indicator. Point estimates are based on bias adjustment using the misclassification rates from the modern replications and GMM using the proxy variable. The lower bound is based on IV. HAC-robust standard errors are given in parentheses. Coefficients with superscripts ***,**,* are statistically significant at the 1%, 5%, 10% level.

proxy, which amounts to $Bias_x = -P[d_t = 1|x_t = 0] - P[d_t = 0|x_t = 1]$.[24] The implied bias for the CPI is significantly different from zero in all specifications and broadly consistent with the misclassification factors based on modern data (see Table 4). The bias ranges from -0.37 to -0.51. For the proxy, the bias is only statistically significantly different from zero at the 5% level when using the GDP output gap. The reason is that the estimates are relatively imprecise. In fact, a Wald-test of whether the bias implied by the two variables is equal cannot be rejected at any conventional significance level. Therefore, the data do not provide evidence in favour of the hypothesis that the proxy based on wholesale prices is more accurately measured.

Table 4 — Point estimates of bias

	GDP Growth	GDP gap	IP growth	IP gap
Bias CPI	-0.39**	-0.39***	-0.51**	-0.37***
	(0.18)	(0.12)	(0.21)	(0.13)
Bias proxy	-0.25	-0.29**	-0.03	-0.26*
	(0.21)	(0.13)	(0.35)	(0.14)
Method	GMM	GMM	GMM	GMM
Observations	100	100	100	100
$Bias_x = Bias_z$ (p-value)	0.69	0.60	0.38	0.64

Note: Point estimates for $Bias_x = -P[d_t = 1|x_t = 0] - P[d_t = 0|x_t = 1]$. Standard errors are based on the Delta method. HAC-robust standard errors are given in parentheses. Coefficients with superscripts ***,**,* are statistically significant at the 1%, 5%, 10% level.

The literature has suggested that asset price declines are more strongly related with real activity shortfalls than CPI deflations (see e.g. Borio et al. 2015). The model can be extended to the case of additional binary covariates. But, the extension hinges on the additional assumptions that the covariates are accurately measured, as well as, that they are conditionally independent of the two proxies. Table 5 includes dummies to control for equity price deflations as well as banking crises.[25] For the model estimated using GMM the table gives the p-value for the J-statistic according to Hansen (1982). The over-identifying restrictions are not rejected in any of the specifications.

Controlling for additional covariates leaves the association between deflation and real activity intact. Both, the bounds as well as the GMM point estimates, suggest that CPI deflations were associated with significantly lower GDP growth. Wald-tests for the GMM

[24] See Appendix A for how this bias term depends on the estimated parameters.

[25] The results based on the improved OLS estimates are robust to including both covariates at the same time. GMM becomes infeasible, however, because many sampling fractions are zero in this case.

specifications suggest that CPI deflations were associated with more-serious declines in real activity than banking crises. In addition, we also find a stronger association for deflation than for equity price declines, at least, when using the output gap measures as dependent variables. Meanwhile, there is no significant difference between CPI deflations and equity price declines when using growth rates of the real activity variables.

Table 5 — Comparison with equity price declines and banking crises

(A) Comparison with equity price declines

	GDP Growth		GDP gap		IP growth		IP gap	
Deflation	-2.37***	-2.28***	-5.46***	-5.80***	-4.50***	-4.83***	-9.06***	-10.12***
	(0.74)	(0.72)	(1.16)	(1.26)	(1.28)	(1.28)	(1.99)	(2.03)
Equity	-1.80**	-1.91***	-1.78*	-2.03**	-3.80***	-3.85***	-3.30*	-3.72**
	(0.73)	(0.66)	(0.97)	(0.90)	(1.38)	(1.28)	(1.70)	(1.88)
Method	Improved	GMM	Improved	GMM	Improved	GMM	Improved	GMM
Observations	100	100	100	100	100	100	100	100
$\beta = \delta$ (p-value)	0.53	0.66	0.00	0.00	0.69	0.55	0.01	0.01
J-test (p-value)		0.34		0.29		0.73		0.43

(B) Comparison with banking crises

	GDP Growth		GDP gap		IP growth		IP gap	
Deflation	-2.18***	-2.89***	-5.29***	-6.97***	-4.13***	-4.95***	-8.77***	-10.94***
	(0.76)	(0.79)	(1.16)	(1.30)	(1.35)	(1.52)	(2.02)	(2.27)
Crises	-2.37*	0.30	-1.77	2.24**	-3.12	4.26**	-0.83	7.52***
	(1.42)	(0.85)	(1.96)	(1.14)	(3.10)	(1.85)	(3.71)	(1.92)
Method	Improved	GMM	Improved	GMM	Improved	GMM	Improved	GMM
Observations	100	100	100	100	100	100	100	100
$\beta = \delta$ (p-value)	0.90	0.02	0.11	0.00	0.76	0.00	0.04	0.00
J-test (p-value)		0.25		0.33		0.46		0.45

Note: Estimates of the model $y_t = \alpha + \beta x_t + \delta q_t + \epsilon_t$. The upper bounds of the association are improved OLS estimates by Black et al. (2000) using the proxy variable as a second indicator. Point estimates are based on GMM. For GMM, the table reports the p-value for the J-statistic by Hansen (1982). HAC-robust standard errors are given in parentheses. Coefficients with superscripts ***,**,* are statistically significant at the 1%, 5%, 10% level.

So far, the improved OLS estimates were relatively close to the GMM point estimates making them quite informative. Therefore, we can estimate more complicated models that would be in principle identified but are difficult to estimate using GMM because of the relatively low number of observations. Panel (A) in Table 6 provides estimates separated into mild and severe deflations using OLS and the improved OLS estimates by Black et al. (2000).[26] A severe deflation is defined as a decline in the CPI and proxy of at

[26]Note that all specifications control for equity price deflations and banking crises but the coefficients are not reported for brevity.

least 3%. The improved OLS estimates measure the shortfall in real activity when both indicators agree that a period was associated with a severe or mild deflation, respectively. The coefficients on mild deflations are not statistically significantly different than zero in most specifications. Meanwhile, severe deflations are associated with significantly lower real activity. Moreover, the bounds are more strongly negative than the OLS estimates suggesting that the misclassification particularly affects severe deflations. This is in line with econometric theory, which implies that the misclassification bias is more severe when the threshold value differs from the unconditional mean of inflation.

In addition, we can examine whether short-lived deflations have a different association with real activity than long-lived ones. Panel (B) splits up deflations into transitory and persistent deflations. A persistent deflation is defined as lasting two years or longer. Focusing on the OLS estimate for GDP growth, we find that there is no significant association for transitory deflations, whereas persistent deflations go hand in hand with shortfalls in economic activity. The bound tells a different story. Both, persistent and transitory deflations are associated with significantly lower GDP growth and the coefficients do not significantly differ from each other. Therefore, the evidence suggests that severe deflations are associated with lower real activity independently of the length of the deflationary episode. Over all specifications, there is little evidence that transitory deflations differ significantly from persistent deflations at least when controlling for measurement error.

The last Panel of Table 6 considers whether monetary deflations have a different association than nonmonetary deflations. A monetary deflation is defined as a decline in the price level that is associated with particularly low growth in M2 of less than 3%.[27] Focusing on GDP and instustrial production growth, neither monetary nor nonmonetary deflations are significantly associated with lower real activity (note that those results are based on a smaller sample from 1868-1899). The bounds, however, indicate that monetary deflations, at the 1% level, as well as other deflations, at the 10% level, are associated with significantly lower GDP growth. For the output gaps, there is no significant difference between the two kinds of deflation as well, and the coefficients become more negative

[27]The M2 series stems from Friedman and Schwartz (1963) as reported by Anderson (2003). The specification hinges on the additional assumption that M2 is measured without error.

Table 6 — Bounds for severe, persistent, and monetary deflations

(A) Severe deflations of more than −3%	GDP Growth		GDP gap		IP growth		IP gap	
Mild	0.09	2.16*	-1.65	1.91	0.07	-0.70	-5.90**	-3.17
	(1.04)	(1.26)	(1.33)	(2.04)	(1.89)	(3.57)	(2.43)	(4.77)
Severe	-3.01***	-3.56***	-5.45***	-7.17***	-4.57***	-5.95***	-6.74***	-9.66***
	(0.67)	(0.72)	(1.13)	(1.27)	(1.42)	(1.49)	(2.25)	(2.34)
Method	OLS	Improved	OLS	Improved	OLS	Improved	OLS	Improved
Observations	100	100	100	100	100	100	100	100
$\beta_1 = \beta_2$ (p-value)	0.00	0.00	0.00	0.00	0.01	0.17	0.75	0.22

(B) Persistent deflations longer than one year	GDP Growth		GDP gap		IP growth		IP gap	
Transitory	-1.31	-2.83***	-4.35**	-4.99***	-3.18	-8.66**	-10.27***	-13.56***
	(1.00)	(0.76)	(1.71)	(1.81)	(2.31)	(3.37)	(2.53)	(5.15)
Persistent	-1.72**	-2.30***	-3.63***	-5.24***	-2.35	-4.32***	-5.40**	-8.86***
	(0.77)	(0.77)	(1.12)	(1.19)	(1.48)	(1.31)	(2.16)	(2.03)
Method	OLS	Improved	OLS	Improved	OLS	Improved	OLS	Improved
Observations	100	100	100	100	100	100	100	100
$\beta_1 = \beta_2$ (p-value)	0.70	0.49	0.67	0.88	0.72	0.17	0.07	0.35

(C) Monetary deflations with M2 growth lower than 3% (1868-1899)	GDP Growth		GDP gap		IP growth		IP gap	
Other	-2.36	-3.02*	-8.62***	-10.53***	-3.54	-5.73*	-12.13***	-16.24***
	(1.67)	(1.74)	(2.21)	(1.46)	(2.75)	(2.93)	(3.11)	(2.27)
Monetary	-1.83	-5.30***	-7.15**	-6.02**	-3.85	-10.04***	-12.72***	-15.88***
	(2.83)	(1.68)	(3.13)	(2.97)	(4.44)	(1.97)	(3.67)	(3.10)
M2	-3.91	-0.37	-3.00	-6.03***	-4.89	0.68	-4.98	-5.01*
	(2.50)	(1.23)	(2.26)	(1.55)	(4.18)	(1.82)	(3.86)	(2.73)
Method	OLS	Improved	OLS	Improved	OLS	Improved	OLS	Improved
Observations	100	100	100	100	100	100	100	100
$\beta_1 = \beta_2$ (p-value)	0.86	0.37	0.71	0.19	0.96	0.25	0.90	0.91

Note: Estimates of the models $y_t = \alpha + \beta_1 x_t^1 + \beta_2 x_t^2 + \delta q_t + \epsilon_t$ using OLS, as well as the improved bound by Black et al. (2000). x_t^1 and x_t^2 denote deflations separated into mild and severe, transitory and persistent, as well as monetary and other deflations. All specifications include equity price deflations as well as banking crises dummies as controls. Those coefficients are not reported for brevity. HAC-robust standard errors are given in parentheses. Coefficients with superscripts ***,**,* are statistically significant at the 1%, 5%, 10% level.

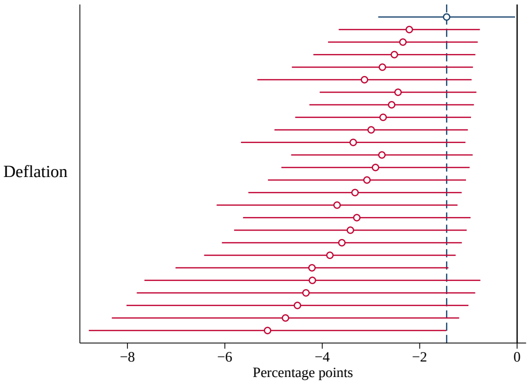
when controlling for measurement error.

So far, we have made extensive use of the assumption that the indicators based on the CPI inflation rate and the proxy are independent conditional on the true deflation indicator. We have seen, however, that the measurement errors of the two inflation measures are positively correlated in modern data. Relaxing this assumption we can still set-identify the parameter of interest. The identifying assumption is that the measurement error in the two indicators is not too severe so that the joint probabilities of misclassfiying inflations ($P[x_t = 1, z_t = 1|d_t = 0]$) and deflations ($P[x_t = 0, z_t = 0|d_t = 1]$) is small. This assumption is supported by the modern replications. The sample analogues of these probabilities amount to 0.05 (misclassified deflation) and 0.15 (misclassified inflation). I therefore estimate the model over a grid of all possible combinations of the two probabilities from 0 to 0.2 in steps of 0.05.[28]

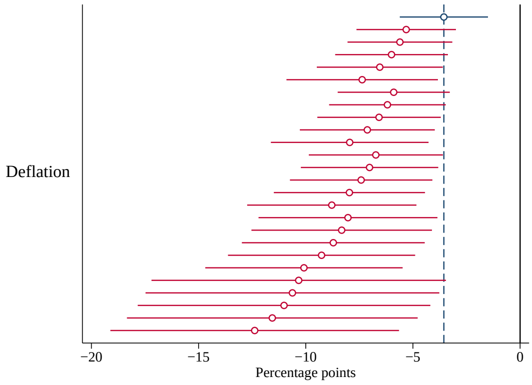
Each panel of Figure 3 shows estimates of the real activity shortfall during deflation. The estimates do not include covariates and therefore correspond to the estimates presented in Table 3. Point estimates are displayed as circles where the horizontal lines represent 95% confidence intervals. The first estimate in each panel, marked by the dashed line, shows the OLS estimate based on the CPI inflation indicator. The remaining estimates are based on GMM at fixed misclassification probabilities. We see that the GMM point estimates are all higher than the OLS point estimates and statistically significantly different from zero. The differences are more pronounced for the output gaps and industrial production growth than for GDP growth. For the former three real activity measures, the GMM estimates are often statistically significantly different from the value of the OLS point estimate. Although the confidence intervals are wide in some cases, it is worth pointing out that accounting for measurement error, the point estimates using GMM are up to three times larger than the OLS estimates, which underlines that descriptive statistics based on erroneous classifications can be substantially biased.

[28]Simulations suggest that, if the measurement error in the two proxies is not too severe (meaning that the signal to noise ratio is unity or larger), the range encompasses a correlation between the measurement error of the two indicators up to 0.9. For a signal to noise ratio of only 0.5, the range of values encompasses a correlation between the errors of 0.7.

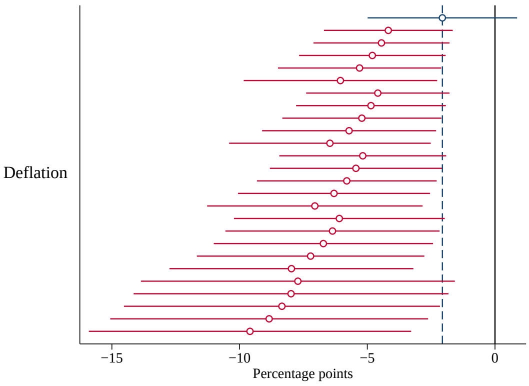
Figure 3 — Relaxing conditional independence



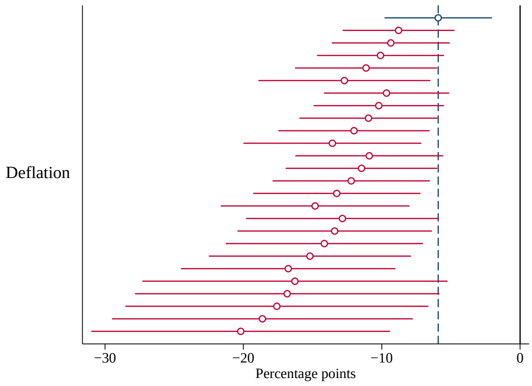
(A) GDP growth



(B) GDP output gap



(C) IP growth



(D) IP output gap

Note: OLS point estimate (blue, dashed vertical line) and GMM estimates (red) on the shortfall of growth during deflations. Circles display point estimates and the solid lines 95% confidence bands based on HAC-robust standard errors. GMM estimates conditional on fixing the joint misclassification rates of inflationary and deflationary periods on a grid of from 0 to 0.2 in steps of 0.05.

4.1 Robustness

The results are robust with respect to various alternative specifications (see Appendix C). In the baseline specification using GDP growth, I excluded the period from 1890-1899 because some of the series underlying the proxy were also used to construct the CPI segment by Rees (1961). I also additionally controlled for severe inflations, and varied the definition of severe deflations (at least -5%), persistent deflations (3 years or longer), as well as monetary deflations (M2 growth lower than 2%). The results are similar. Using HP-filtered output gaps instead of the procedure of Hamilton (2016) is inconsequential as well. In addition, I examined whether using the real per capita GDP growth from The Maddison-Project (2013), which are largely based on Sutch (2006), changes the result. OLS implies a weak association with an insignificant 0.6 percentage points drop in GDP growth. The improved bound is significantly different from zero at the 10% level and amounts to -1.2 percentage points. Therefore, while the estimation uncertainty is larger, the implied bias is more severe.

Then, I changed the deflation threshold to 1%, taking into account for the possibility that similar measurement issues as emphasized by the Boskin Commission (1996) affects historical inflation data. Qualitatively, the results are similar although less precisely estimated. It turns out that individual estimates using the CPI deflation indicator are not significant. The improved upper bound suggests that deflation was associated at least with 1.8 percentage points drop in GDP growth. The GMM point estimate is significant only at the 10% level and amounts to -1.6 percentage points. The lower bound based on IV, however, suggests that the GDP growth shortfall amounts up to 4.4 percentage points.

5 Conclusions

Estimating average real economic performance during deflations is hampered by measurement error. Replications of deficiencies in 19th century CPI estimates suggest that those measurement issues bias the link between real economic activity and deflation. Using four different approaches to alleviate the errors-in-variables problem I find that deflations were associated with substantially lower real activity for the 19th century US. Moreover, perhaps surprisingly, transitory deflations were associated with similar shortfalls in real activity as more persistent ones. The association, however, is limited to severe deflations.

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. This paper argues that measurement problems in historical price data is partly responsible for the lacking association.

Whether deflation causes lower real activity or whether it is a consequence of falling aggregate demand remains an open question. Accurately estimating reduced-form correlations, however, is a necessary condition for reliable structural analysis. Most estimation approaches to identify the impact of structural shocks will suffer from the errors-in-variables problem. Exploring 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

AIGNER, D. J. (1973): "Regression With a Binary Independent Variable Subject to Errors of Observation," *Journal of Econometrics*, 1, 49–60.

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

ANDERSON, R. G. (2003): "Some Tables of Historical U.S. Currency and Monetary Aggregates Data," Working Papers 2003-006, Federal Reserve Bank of St. Louis.

ATKESON, A. AND P. J. KEHOE (2004): "Deflation and Depression: Is There an Empirical Link?" *American Economic Review*, 94, 99–103.

BAUM, C. F. AND M. E. SCHAFFER (2013): "ACTEST: Stata module to perform Cumby-Huizinga general test for autocorrelation in time series," Statistical Software Components, Boston College Department of Economics.

BAYOUMI, T. AND B. EICHENGREEN (1996): "The Stability of the Gold Standard and the Evolution of the International Monetary Fund," in *Modern Perspectives on the Gold Standard*, ed. by T. Bayoumi, B. Eichengreen, and M. P. Taylor, Cambridge: Cambridge University Press, chap. 6, 165–188.

BECKWORTH, D. (2007): "The Postbellum Deflation and its Lessons for Today," *The North American Journal of Economics and Finance*, 18, 195–214.

BERNHOLZ, P. (2003): *Monetary Regimes and Inflation*, Cheltenham: Edward Elgar Publishing.

BLACK, D. A., M. C. BERGER, AND F. A. SCOTT (2000): "Bounding Parameter Estimates with Nonclassical Measurement Error," *Journal of the American Statistical Association*, 95, 739–748.

BLS (1923): "Retail Prices 1913 to December, 1921," Bulletin 315, U.S. Bureau of Labor Statistics.

BORDO, M. 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 A. REDISH (2004): "Is Deflation Depressing? Evidence From the Classical Gold Standard," in *Deflation: Current and Historical Perspectives*, ed. by R. Burdekin and P. Siklos, Cambridge: Cambridge University Press, chap. 7, 99–202.

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 `www.ssa.gov/history/reports/boskinrpt.html`.

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, 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 490.

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

Friedman, M. and A. Schwartz (1963): *A Monetary History of the United States, 1867–1960*, Princeton: Princeton University Press.

Gordon, R. J. (2016): *The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War*, Princeton: Princeton University Press.

Griliches, Z. (1986): "Economic Data Issues," in *Handbook of Econometrics*, ed. by Z. Griliches and M. D. Intriligator, Amsterdam: Elsevier, vol. 3, chap. 25, 1465–1514.

Hamilton, J. D. (2016): "Why you Should Never Use the Hodrick-Prescott Filter," retrieved from `econweb.ucsd.edu/~jhamilto/hp.pdf`, Department of Economics, UC San Diego.

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

——— (2006): "Wholesale Price Indexes, by Commodity Group: 1749–1890 [Warren Pearson], Table Cc113-124," in *Historical Statistics of the United States, Earliest Times to the Present: Millennial Edition*, ed. by S. B. Carter, S. S. Gartner, M. R. Haines, A. L. Olmstead, R. Sutch, and G. Wright, New York: Cambridge University Press.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

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

Hoover, E. D. (1960): "Retail Prices after 1850," in *Trends in the American Economy in the Nineteenth Century*, National Bureau of Economic Research, 141–190.

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

Johnston, L. and S. H. Williamson (2008): "The Annual Real and Nominal GDP for the United States, 1790-1928," retrieved from `www.measuringworth.com/usgdp/`, MeasuringWorth.

——— (2016): "What Was the U.S. GDP Then?" retrieved from `www.measuringworth.com/usgdp/`, MeasuringWorth.

Jorda, O., M. Schularick, and A. M. Taylor (2016): "Macrofinancial History and the New Business Cycle Facts," in *NBER Macroeconomics Annual 2016*, National Bureau of Economic Research, vol. 31.

KANE, T. J., C. E. ROUSE, AND D. STAIGER (1999): "Estimating Returns to Schooling When Schooling is Misreported," NBER Working Papers 7235, National Bureau of Economic Research.

KAUFMANN, D. (2016): "Is Deflation Costly After All? Evidence from Noisy Historical Data," Working Papers 421, KOF Swiss Economic Institute.

——— (2017): "Measurement Error in 19th Century CPI Data. Evidence from Modern Replications," mimeo, KOF Swiss Economic Institute, ETH Zurich.

KNOLL, K., M. SCHULARICK, AND T. STEGER (2017): "No Price Like Home: Global House Prices, 1870-2012," *American Economic Review*, 107, 331–53.

KOMUNJER, I. AND S. NG (2014): "Measurement Errors in Dynamic Models," *Econometric Theory*, 30, 150–175.

KREIDER, B. (2010): "Regression Coefficient Identification Decay in the Presence of Infrequent Classification Errors," *The Review of Economics and Statistics*, 92, 1017–1023.

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

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

MUMTAZ, H., G. PINTER, AND K. THEODORIDIS (2015): "What do VARs Tell us About the Impact of a Credit Supply Shock?" Working Paper 739, Queen Mary University of London.

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, 268–281.

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

OFFICER, L. H. AND S. H. WILLIAMSON (2016): "Annual Consumer Price Index for the United States, 1774-2015," retrieved from `www.measuringworth.com/uscpi/`, MeasuringWorth.

PAKES, A. (1982): "On the Asymptotic Bias of Wald-Type Estimators of a Straight Line when Both Variables are Subject to Error," *International Economic Review*, 23, 491–497.

REES, A. (1961): *Real Wages in Manufacturing, 1890-1914*, Princeton: Princeton University Press.

REINHART, C. M. AND K. S. ROGOFF (2010): "Growth in a Time of Debt," *American Economic Review*, 100, 573–578.

ROMER, C. (1986a): "Is the Stabilization of the Postwar Economy a Figment of the Data?" *American Economic Review*, 76, 314–3347.

——— (1986b): "Spurious Volatility in Historical Unemployment Data," *Journal of Political Economy*, 94, 1–37.

Schwert, G. W. (1989): "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business & Economic Statistics*, 7, 147–159.

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

Staudenmayer, J. and J. P. Buonaccorsi (2005): "Measurement Error in Linear Autoregressive Models," *Journal of the American Statistical Association*, 100, 841–852.

Sutch, R. (2006): "Historical Statistics of the United States, Earliest Times to the Present: Millennial Edition," New York: Cambridge University Press, chap. National Income and Product.

The Maddison-Project (2013): "2013 version," retrieved from www.ggdc.net/maddison/maddison-project/, The Maddison-Project.

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

Appendices

A Technical appendix

This section derives some analytic results for comparing the attenuation and misclassification biases and discusses the GMM estimation approach in detail.

A.1 Attenuation and misclassification bias

Let y_t denote the dependent variable, π_t the correctly measured variable (signal), and ω_t the measurement error (noise). In addition, let the proxy of inflation be denoted by $\tilde{\pi}_t = \rho_0 + \rho_1 \pi_t + \omega_t$ and the error-ridden deflation dummy as $x_t \equiv \mathbf{1}_{\{\tilde{\pi}_t < c\}}$. Finally, let ε_t denote a stochastic disturbance. All variables, except y_t, are normally distributed and mutually independent. Note that, for simplicity, I assume that there is no mismeasurement in the mean and scale of inflation ($\rho_0 = 0,\ \rho_1 = 1$).

It is well known that classical measurement error in a continuous variable in the regression

$$y_t = \alpha + \beta_c \tilde{\pi}_t + \varepsilon_t$$

implies that the OLS estimate is biased and depends on the relative amount of measurement error in the observed variable (see e.g. Griliches 1986):

$$plim\ \hat{\beta}_c = \beta_c \frac{\sigma_\pi^2}{\sigma_\pi^2 + \sigma_\omega^2}\ ,$$

where the estimate is attenuated towards zero because variances are strictly positive. It follows that the attenuation bias only depends on the signal to noise ratio but not on mismeasurement of the mean.

Aigner (1973) showed that in the corresponding binary model

$$y_t = \alpha + \beta_b x_t + \varepsilon_t\ ,$$

OLS is inconsistent as well with

$$plim\ \hat{\beta}_b = \beta_b (1 - P[d_t = 0 | x_t = 1] - P[d_t = 1 | x_t = 0])\ ,$$

where $d_t \equiv \mathbf{1}_{\{\pi_t < c\}}$ is the well-measured deflation dummy. Because probabilities lie between zero and unity, the OLS coefficient is generally biased upwards (downwards) if β is negative (positive).

To derive some analytical insights into the differences of the bias in the binary relative to the continuous case let us assume that the signal and the noise variable both follow a normal distribution with the same standard deviation σ. This implies that the signal to noise ratio is unity. In the continuous case, the attenuation factor therefore amounts to 0.5.

In addition, let us assume that both variables are of zero mean. A nonzero threshold therefore implies that it differs from the mean of the signal. Because of the independence assumption it follows that we can write the second term of the misclassification bias as:

$$P[\pi_t < c | \pi_t + \omega_t > c] = P[\max(\pi_t, \omega_t) < c] = P[\pi_t < c] P[\omega_t < c] \ ,$$

and because of normality we obtain

$$P[\pi_t < c | \pi_t + \omega_t > c] = \Phi\left(\frac{c}{\sigma}\right)^2 \ ,$$

where $\Phi(z)$ is the cdf of a standard normal distribution.

Similarly, we obtain for the first term that:

$$P[\pi_t > c | \pi_t + \omega_t < c] = P[-\pi_t < -c | -\pi_t - \omega_t > -c]$$

because the random variables are assumed to have zero mean and are symmetrically distributed it follows that

$$P[\pi_t > c | \pi_t + \omega_t < c] = \Phi\left(\frac{-c}{\sigma}\right)^2$$

Therefore, under these admittedly strong assumptions, we can derive that the misclassification factor for the binary variable amounts to:

$$1 - P[d_t = 0 | x_t = 1] - P[d_t = 1 | x_t = 0] = 1 - \Phi\left(\frac{c}{\sigma}\right)^2 - \Phi\left(\frac{-c}{\sigma}\right)^2$$

Some interesting observations emerge. In contrast to the continuous case the bias now depends on the threshold value, that is, the deviation of the threshold from the mean of the signal. If this deviation is zero ($c = 0$), the bias amounts to 0.5, which is equal to the bias in the continuous case. If the threshold is different from the mean of the signal ($c \neq 0$), however, the bias becomes more severe.

In addition, the overall volatility, not only the signal to noise ratio, matters for the bias. If the overall variance of the process, which depends on σ, is large, then the bias will be less severe than if the overall variance is small. This suggests that the misclassification bias will be less severe if the signal is stronger, even if the noise process gets proportionally more volatile. This suggests that the misclassification bias may be particularly severe for variables with little variation, whereas, in the continuous case the bias does not depend on the overall volatility but only on the signal to noise ratio.

A.2 GMM estimators with mismeasured binary regressors

Resolutions to the misclassification bias differ from the well-known attenuation bias in the case of classical measurement error. With classical measurement error, a widely used solution is to find a proxy variable with the same properties as the error-ridden variable, but with independent measurement error, and use it as an instrument (see Hausman 2001). It is worth noting, however, that the classical assumptions are violated in our case because the classification error is necessarily negatively correlated with the outcome (see Kane et al. 1999). Following Kane et al. (1999) and Black et al. (2000) this appendix derives a consistent GMM estimator for the regression model with an error-ridden binary regressor and two extensions.

A.2.1 The basic model

Suppose that the true model reads

$$y_t = \alpha + \beta d_t + \varepsilon_t \ ,$$

with $d_t \equiv \mathbf{1}_{\{\pi_t<0\}}$ but we have at our disposal only two error-ridden indicators $x_t \equiv \mathbf{1}_{\{\tilde{\pi}_t<0\}}$, where $\tilde{\pi}_t = \rho_0 + \rho_1 \pi_t + \omega_t$, and $z_t \equiv \mathbf{1}_{\{\hat{\pi}_t<0\}}$, where $\hat{\pi}_t = \gamma_0 + \gamma_0 \pi_t + \psi_t$. Assume that $\psi_t, \omega_t, \varepsilon_t$ are *i.i.d.* and mutually independent of each other and of π_t.

Based on information of only one of the indicators, the model is not identified and we will not be able to recover the coefficients. For example, using only the information in x_t we can estimate three independent moments from the data, namely the expectation of y_t conditional on each outcome of x_t and the population probability of $x_t = 1$:

$$\begin{aligned} E[y_t|x_t = 1] &= \alpha + \beta P[d_t = 1|x_t = 1] \\ E[y_t|x_t = 0] &= \alpha + \beta P[d_t = 1|x_t = 0] \\ P[x_t] &= P[x_t = 1|d_t = 1]P[d_t = 1] + P[x_t = 1|d_t = 0]P[d_t = 0] \\ &= (1 - P[x_t = 0|d_t = 1])P[d_t = 1] + P[x_t = 1|d_t = 0](1 - P[d_t = 1]) \ . \end{aligned}$$

We can apply Bayes' theorem to rewrite the conditional probabilities in the conditional expectation as

$$\begin{aligned} P[d_t = 1|x_t = 1] &= \frac{P[x_t = 1|d_t = 1]P[d_t = 1]}{P[x_t]} \\ &= \frac{(1 - P[x_t = 0|d_t = 1])P[d_t = 1]}{(1 - P[x_t = 0|d_t = 1])P[d_t = 1] + P[x_t = 1|d_t = 0]P[d_t = 0]} \ . \end{aligned}$$

Let us define the population probability of deflation as $p \equiv P[d_t = 1]$, as well as the probability that a deflationary and inflationary episode is misclassified by the indicator as $\eta_x \equiv P[x_t = 0|d_t = 1]$, $\nu_x \equiv P[x_t = 1|d_t = 0]$. Then, we can rewrite the moment conditions as

$$\begin{aligned} E[y_t|x_t = 1] &= \alpha + \beta \frac{(1 - \eta_x)p}{(1 - \eta_x)p + \nu_x(1 - p)} \\ E[y_t|x_t = 0] &= \alpha + \beta \frac{\eta_x p}{\eta_x p + (1 - \nu_x)(1 - p)} \\ P[x_t] &= (1 - \eta_x)p + \nu_x(1 - p) \ . \end{aligned}$$

We see that we have to estimate five coefficients from three moment conditions. Thus the model is not identified. It is easy to show that OLS is not consistent:

$$\begin{aligned} plim\ \hat{\beta}_{OLS} &= E[y_t|x_t = 1] - E[y_t|x_t = 0] \\ &= \beta\left(\frac{(1-\eta_x)p}{(1-\eta_x)p + v_x(1-p)} - \frac{\eta_x p}{\eta_x p + (1-\nu_x)(1-p)}\right) \ . \end{aligned}$$

However, using the information of the two binary indicators it is possible to consistently estimate β using GMM (see Kane et al. 1999, Black et al. 2000). To see this, note that we can derive the expected value of y_t conditional on all combinations of outcomes of the two indicators which yields four moments. In addition, we can derive the probability of each of the four combinations of outcomes which yields another three independent moments.[29] For example, for the case of $x_t = 1$ and $z_t = 1$ we have:

$$E[y_t|x_t = 1, z_t = 1] = \alpha + \beta P[d_t = 1|x_t = 1, z_t = 1] \ .$$

Applying Bayes' theorem we can rewrite the conditional probability as

$$P[d_t = 1|x_t = 1, z_t = 1] = \frac{P[x_t = 1, z_t = 1|d_t = 1]P[d_t = 1]}{P[x_t = 1, z_t = 1]} \ .$$

Using the assumption that x_t and z_t are independent conditional on the actual outcome of d_t we have

$$\begin{aligned} P[d_t = 1|x_t = 1, z_t = 1] &= \frac{P[x_t = 1|d_t = 1]P[z_t = 1|d_t = 1]P[d_t = 1]}{P[x_t = 1, z_t = 1]} \qquad \text{(A.1)} \\ &= \frac{(1 - P[x_t = 0|d_t = 1])(1 - P[z_t = 0|d_t = 1])P[d_t = 1]}{P[x_t = 1, z_t = 1]} \ . \end{aligned}$$

The same strategy can be applied to rewrite the denominator as:

$$\begin{aligned} P[x_t = 1, z_t = 1] &= P[x_t = 1, z_t = 1|d_t = 1]P[d_t = 1] \qquad \text{(A.2)} \\ &\quad + P[x_t = 1, z_t = 1|d_t = 0]P[d_t = 0] \\ &= (1 - P[x_t = 0|d_t = 1])(1 - P[z_t = 0|d_t = 1])P[d_t = 1] \\ &\quad + P[x_t = 1|d_t = 0]P[z_t = 1|d_t = 0](1 - P[d_t = 1]) \ . \end{aligned}$$

[29] One probability is redundant because the joint probabilities over all combinations of outcomes of x_t and z_t have to sum to unity.

Using equation (A.2) in equation (A.1) and previously introduced notation yields expressions for two of the seven moments:

$$
\begin{aligned}
E[y_t|x_t = 1, z_t = 1] &= \alpha + \beta \frac{(1-\eta_x)(1-\eta_z)p}{(1-\eta_x)(1-\eta_z)p + \nu_x\nu_z(1-p)} \\
P[x_t = 1, z_t = 1] &= (1-\eta_x)(1-\eta_z)p + \nu_x\nu_z(1-p) \ .
\end{aligned}
$$

We can derive similar expressions for the remaining three conditional expectations and two non-redundant probabilities in terms of the misclassifcation probabilities (η_x, ν_x, η_z, ν_z), model parameters (α, β) and the population probability of deflation p. Therefore we have seven parameters to estimate and seven moments that we can estimate from the data. Generally, if we obtain M noisy binary indicators this yields $2 \times 2^M - 1$ moment conditions and $3 + 2 \times M$ coefficients to estimate. So with two indicators, the model is just identified and with more indicators, the model is over-identified.

GMM allows us to estimate various measures of interest. In particular, we can derive a point estimate of the implied bias associated with the two indicators. Given the discussion so far, it is easy to show that the for indicator x_t amounts to

$$
\begin{aligned}
Bias_x &= -P[x_t = 0|d_t = 1] - P[x_t = 1|d_t = 0] \\
&= -\frac{\eta_x p}{\eta_x p + (1-\nu_x)(1-p)} - \frac{\nu_x(1-p)}{(1-\eta_x)p + \nu_x(1-p)} \ .
\end{aligned}
$$

An analogous formula applies to indicator z_t. Given the the parameters estimates based on GMM, we can derive standard errors for the bias using the Delta method.

A.2.2 Well-measured binary covariates

The estimator can be extended to including binary covariates. Let us assume that we have one additional binary covariate and estimate the model:

$$
y_t = \alpha + \beta x_t + \delta q_t + \epsilon_t \ .
$$

Focusing on the case with $x_t = 1$, $z_t = 1$ and $q_t = 1$ we have:

$$
\begin{aligned}
E[y_t|x_t = 1, z_t = 1, q_t = 1] &= \alpha + \beta P[d_t = 1|x_t = 1, z_t = 1, q_t = 1] \\
&\quad + \delta P[q_t = 1|x_t = 1, z_t = 1, q_t = 1] \ .
\end{aligned}
$$

Assuming that the covariate is accurately measured implies $P[q_t = 1|x_t = 1, z_t = 1, q_t = 1] = 1$. Therefore, can proceed analogous to the previous case applying Bayes' theorem:

$$
P[d_t = 1|x_t = 1, z_t = 1, q_t = 1] = \frac{P[x_t = 1, z_t = 1, q_t = 1|d_t = 1]P[d_t = 1]}{P[x_t = 1, z_t = 1, q_t = 1]} \ .
$$

Let us again assume that the deflation indicators are conditionally independent of each other and of the additional covariate to obtain:

$$
P[d_t = 1|x_t = 1, z_t = 1, q_t = 1] = \frac{(1-\eta_x)(1-\eta_z)(1-\eta_q)p}{(1-\eta_x)(1-\eta_z)(1-\eta_q)p + \nu_x\nu_z\nu_q(1-p)} \ .
$$

Therefore we have

$$
E[y_t|x_t = 1, z_t = 1, q_t = 1] = \alpha + \beta\frac{(1-\eta_x)(1-\eta_z)(1-\eta_q)p}{(1-\eta_x)(1-\eta_z)(1-\eta_q)p + \nu_x\nu_z\nu_q(1-p)} + \delta \ ,
$$

where the corresponding population moment for the sampling fraction, $P[x_t = 1, z_t = 1, q_t = 1]$, equals the denominator. Again, we can derive analogous expressions for all combinations of outcomes, which yields $2 \times 2^{M+1} - 1 = 15$ with $M = 2$ indicators. However, we only have to estimate $3 + 1 + 2 \times (M + 1) = 10$ coefficients implying that, if we include an accurately measured binary covariate, the model is over-identified.

The estimator readily extends to more than one binary covariate, categorical outcomes, as well as deterministic time-period interaction terms. In these extensions the model is always just or over-identified, if we have two conditionally independent deflation indicators.

A.2.3 Relaxing conditional independence

Without the conditional independence assumption, we cannot write the joint conditional probabilities as separate parameters. As a consequence, we have for the case that $x_t = 1, z_t = 1$:

$$E[y_t|x_t = 1, z_t = 1] \quad = \quad \alpha + \beta \frac{\eta_{11} p}{\eta_{11} p + \nu_{11}(1-p)} \ . \tag{A.3}$$

where $\eta_{11} \equiv P[x_t = 1, z_t = 1|d_t = 1]$ and $\nu_{11} \equiv P[x_t = 1, z_t = 1|d_t = 0]$. Conditioning on every combination of outcomes shows that we can still estimate seven independent moments from the data. However, we have to estimate nine coefficients: the two model parameters, the probability of deflation, and three joint conditional probabilities for every outcome of the true deflation indicator $(\eta_{11}, \eta_{10}, \eta_{00}, \nu_{11}, \nu_{10}, \nu_{00})$.

Two of the parameters to be estimated are the conditional probability that both indicators misclassify a deflationary period and inflationary period, respectively, at the same time:

$$\eta_{00} = P[x_t = 0, z_t = 0|d_t = 1], \ \nu_{11} = P[x_t = 1, z_t = 1|d_t = 0] \ .$$

If we fix those probabilities at sensible values, we only have to estimate 7 parameters and the model is identified.

The two parameters are particularly useful to fix because, if we think that the two indicators exhibit measurement error that is not too severe and the measurement error is not too strongly correlated, the joint misclassification probabilities should be small. If we can define a reasonable upper threshold for the two joint probabilities the parameter of interest is set-identified.

B Data

Table B.1 — Sources

Name	Time	Source	Identifier	Comments
Composite aggregate price indices				
CPI	1774-2015	MW		Officer and Williamson (2016)
WPI/PPI	1749-1890	HSUS	Cc113	Warren and Pearson (1933)
	1860-1990	HSUS	Cc125	Hanes (1998); 1941-1947 missing. The series is a consistent replication of Warren and Pearson (1933) for the 19th century
	1913-2015	FRED	PPIACO	
Composite real activity				
Real GDP	1790-2015	MW		Johnston and Williamson (2016); per capita
Industrial production	1790-1915			Davis (2004)
	1899-1937	HSUS	Dd495	Fabricant (1940)
	1919-2015	FRED	INDPRO	
Output gaps	1790-2015			Detrended GDP and industrial production using the approach by Hamilton (2016) and a HP-filter with smoothing parameter set to 100
Modern replications				
WPI	1890-1990	HSUS	Cc125	Hanes (1998); 1941-1947 missing. The series is a consistent replication of Warren and Pearson (1933) for the 19th century
CPI	1956-2015			Kaufmann (2017) the series is a replication of the 1860-1880 segment of the CPI provided by Officer and Williamson (2016)
BLS PPI data for proxy				
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	
Historical WPI data for proxy				
Foods	1798-1941	HSUS	Cc115, Cc128	Warren and Pearson (1933) extended by Hanes (1998)
Textile products	1798-1941	HSUS	Cc117, Cc130	
Fuel and lighting	1798-1941	HSUS	Cc118, Cc131	

continued on next page

Table B.1 – *continued from previous page*

Name	Time	Source	Identifier	Comments
House furnishing goods	1840-1941	HSUS	Cc122, Cc135	
Other historical variables				
Banking crises				Jalil (2015); 1833-1834, 1837-1839, 1857, 1873, 1893, 1907
Equity prices	1802-1870 1870-2015	HSUS JST	Cj797	Index of common stocks
Money supply	1867-1947			M2 by Friedman and Schwartz (1963) as reported by Anderson (2003)

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: Jorda et al. (2016) and Knoll et al. (2017).

C Robustness

Table C.2 — Real activity during deflations

(A) GDP growth (1800-1889)

	Upper bound		Point estimate		Lower bound
Deflation	-1.44**	-2.09***	-2.08**	-2.18***	-3.26**
	(0.64)	(0.75)	(0.93)	(0.75)	(1.35)
Method	OLS	Improved	Adjusted	GMM	IV
Observations	100	100	100	100	100

(B) HP-filtered GDP output gap

	Upper bound		Point estimate		Lower bound
Deflation	-2.59***	-3.88***	-3.76***	-4.20***	-6.42***
	(0.71)	(0.72)	(1.03)	(0.79)	(1.60)
Method	OLS	Improved	Adjusted	GMM	IV
Observations	100	100	100	100	100

(C) HP-filtered Industrial production output gap

	Upper bound		Point estimate		Lower bound
Deflation	-3.23**	-5.30***	-4.68**	-5.70***	-9.37***
	(1.29)	(1.36)	(1.87)	(1.39)	(2.69)
Method	OLS	Improved	Adjusted	GMM	IV
Observations	100	100	100	100	100

(D) GDP growth from The Maddison-Project (2013)

	Upper bound		Point estimate		Lower bound
Deflation	-0.58	-1.20*	-0.84	-1.24	-2.44*
	(0.74)	(0.67)	(1.08)	(0.80)	(1.48)
Method	OLS	Improved	Adjusted	GMM	IV
Observations	99	99	99	100	99

(E) Deflation threshold at 1%

	Upper bound		Point estimate		Lower bound
Deflation	-0.65	-1.75**	-0.94	-1.60*	-4.41**
	(0.65)	(0.74)	(0.94)	(0.87)	(1.84)
Method	OLS	Improved	Adjusted	GMM	IV
Observations	100	100	100	100	100

Note: Estimates of the model $y_t = \alpha + \beta x_t + \epsilon_t$. The upper bounds of the association are OLS estimates, as well as, the improved OLS estimates by Black et al. (2000) using the proxy variable as a second indicator. Point estimates are based on bias adjustment using the misclassification rates from the modern replications and GMM using the proxy variable. The lower bound is based on IV. HAC-robust standard errors are given in parentheses. Coefficients with superscripts ***,**,* are statistically significant at the 1%, 5%, 10% level.

Table C.3 — Bounds for severe, persistent, and monetary deflations

	GDP Growth		GDP gap		IP growth		IP gap	
(A) Severe deflations of more than −5%								
Mild	-1.23	0.58	-3.15***	-1.44	-1.83	-1.63	-5.97***	-6.20**
	(0.79)	(0.92)	(1.15)	(1.13)	(1.50)	(1.91)	(2.11)	(2.42)
Severe	-2.61***	-3.51***	-5.26***	-7.48***	-4.16**	-6.52***	-7.32***	-10.58***
	(0.79)	(0.84)	(1.30)	(1.39)	(1.85)	(2.17)	(2.71)	(2.64)
Method	OLS	Improved	OLS	Improved	OLS	Improved	OLS	Improved
Observations	100	100	100	100	100	100	100	100
$\beta_1 = \beta_2$ (p-value)	0.01	0.00	0.09	0.00	0.19	0.06	0.62	0.14

	GDP Growth		GDP gap		IP growth		IP gap	
(B) Severe inflations of more than 3%								
Severe inflation	-1.14	-1.19	-2.37	-2.55*	-2.98	-3.01*	-4.74	-4.83*
	(0.87)	(0.84)	(1.53)	(1.36)	(1.90)	(1.79)	(2.93)	(2.71)
Mild deflation	-0.25	1.79	-2.35	1.11	-0.81	-1.65	-7.30***	-4.68
	(1.10)	(1.30)	(1.43)	(2.14)	(1.92)	(3.70)	(2.62)	(4.99)
Severe deflation	-3.36***	-3.94***	-6.18***	-7.99***	-5.48***	-6.92***	-8.18***	-11.21***
	(0.74)	(0.78)	(1.23)	(1.35)	(1.52)	(1.66)	(2.37)	(2.46)
Method	OLS	Improved	OLS	Improved	OLS	Improved	OLS	Improved
Observations	100	100	100	100	100	100	100	100

	GDP Growth		GDP gap		IP growth		IP gap	
(C) Persistent deflations longer than two years								
Transitory	-1.12	-1.53**	-3.48***	-3.30***	-2.01	-4.88**	-6.00***	-6.84**
	(0.78)	(0.76)	(1.18)	(1.25)	(1.65)	(1.98)	(2.25)	(3.39)
Persistent	-2.00**	-2.69***	-3.97***	-5.87***	-2.86*	-4.91***	-6.62***	-10.00***
	(0.86)	(0.91)	(1.30)	(1.38)	(1.67)	(1.37)	(2.43)	(2.21)
Method	OLS	Improved	OLS	Improved	OLS	Improved	OLS	Improved
Observations	100	100	100	100	100	100	100	100
$\beta_1 = \beta_2$ (p-value)	0.35	0.23	0.72	0.07	0.64	0.99	0.82	0.39

	GDP Growth		GDP gap		IP growth		IP gap	
(D) Monetary deflations with M2 growth lower than 2% (1868-1899)								
Other	-2.55	-3.55**	-8.86***	-10.48***	-4.26	-6.96***	-13.13***	-17.23***
	(1.59)	(1.56)	(2.00)	(1.34)	(2.68)	(2.69)	(2.96)	(2.25)
Monetary	-1.14	-3.10*	-6.17**	-5.92***	-1.44	-4.89	-9.36**	-11.45***
	(2.21)	(1.88)	(2.64)	(2.25)	(3.64)	(3.61)	(3.73)	(3.63)
M2	-4.44***	-2.60	-3.65**	-5.56***	-7.00**	-4.43	-7.94**	-8.89***
	(1.69)	(1.59)	(1.53)	(0.76)	(3.20)	(3.46)	(3.22)	(2.96)
Method	OLS	Improved	OLS	Improved	OLS	Improved	OLS	Improved
Observations	100	100	100	100	100	100	100	100
$\beta_1 = \beta_2$ (p-value)	0.48	0.80	0.34	0.05	0.38	0.54	0.35	0.12

Note: Estimates of the models $y_t = \alpha + \beta_1 x_t^1 + \beta_2 x_t^2 + \delta q_t + \epsilon_t$ using OLS, as well as the improved bound by Black et al. (2000). x_t^1 and x_t^2 denote deflations separated into mild and severe, transitory and persistent, as well as monetary and other deflations. All specifications include equity price deflations as well as banking crises dummies as controls. HAC-robust standard errors are given in parentheses. Coefficients with superscripts ***,**,* are statistically significant at the 1%, 5%, 10% level.