Received: 2 July 2019 | Revised: 7 February 2020
DOI: 10.1002/jae.2762

RESEARCH ARTICLE

WILEY Journal of APPLIED ECONOMETRICS

Is deflation costly after all? The perils of erroneous historical classifications

Daniel Kaufmann

University of Neuchâtel and KOF Swiss Economic Institute, Neuchâtel, Switzerland

Correspondence
Daniel Kaufmann, University of Neuchâtel, Rue A.-L. Breguet 2, CH-2000 Neuchâtel, Switzerland.
Email: daniel.kaufmann@unine.ch

Summary

I estimate average economic activity during periods of inflation and deflation while accounting for measurement errors in 19th century prices. These measurement errors lead to underestimation (overestimation) of economic activity during periods of inflation (deflation). By exploiting multiple deflation indicators, it is possible to recover the true relationship; the shortfall of US industrial production growth during periods of deflation ranges from −4.5 pp to −7.6 pp, instead of −2 pp. I also find a negative relationship between deflation and real activity in the UK. I then examine the cross-country variation in the estimates for eleven countries. The patterns are consistent with stronger biases for countries with more serious measurement errors in prices.

1 | INTRODUCTION

In the wake of the 2008 financial crisis, central bankers justified unconventional policies by the risk of deflation (see, e.g., Bernanke, 2010; Hartmann & Smets, 2018; SNB, 2011). There is substantial disagreement, however, whether deflation is harmful. In fact, many deflationary episodes during the 19th century appear benign (Atkeson & Kehoe, 2004; Bordo & Filardo, 2005; Borio et al., 2015). This paper shows that if we fail to account for measurement errors in retrospectively estimated price indices, these deflationary episodes appear more benign than they actually were.

How costly is deflation? Following Borio et al. (2015) and Eichengreen et al. (2016) I cast this question in a linear regression framework with real activity as the dependent variable and a binary deflation indicator as the independent variable.[1] If this binary indicator is measured with error, OLS is biased, for two reasons. First, OLS suffers from misclassification bias (see Aigner, 1973). If we use a mismeasured price index to classify inflationary (deflationary) episodes, some of these episodes are associated with falling (rising) prices and therefore low (high) real activity. Second, OLS suffers from deflator bias if the measurement errors in the GDP deflator are correlated with the measurement errors in the binary deflation indicator. Some of the inflation (deflation) episodes are associated with positive (negative) measurement errors in the GDP deflator, which lead to negative (positive) measurement errors in real GDP.

I propose two strategies to resolve these biases. First, I use multiple error-ridden binary deflation indicators for the 19th century to resolve the misclassification bias. The estimators build on Kane et al. (1999) and Black et al. (2000), who address misreported binary responses in survey data. Their estimators require measurement errors in the binary indicators to be unrelated.[2] However, most historical price indices suffer from similar measurement errors. I therefore propose an

[1]The regression framework makes it possible to quantitatively assess and resolve the bias. Similar issues, however, arise in qualitative analyses examining economic performance during periods with inflation and deflation (see, e.g., Friedman & Schwartz, 1963, p. 96).
[2]Mahajan (2006) proposes an estimator using two binary indicators allowing for correlation between the measurement errors and other regressors. Similar to Kane et al. (1999), however, the measurement errors in the second indicator have to be unrelated to the measurement errors in the first.

estimator allowing for correlated measurement errors under the assumption that the joint misclassification rates are small. To support this assumption I provide external evidence on the misclassification rates using modern replications of 19th century price indices. Second, to address the deflator bias, I focus the main analysis on real activity data that does not depend on a deflator (US industrial production) and on a country with arguably well-measured 19th century data (the UK).

I find measurement errors in historical price data bias the empirical link between real activity and deflation. Deflation is on average associated with a 2.1 pp decline in US industrial production growth if we ignore measurement errors. If we use a second binary deflation indicator the shortfall amounts to −4.5 pp. If we allow for correlated measurement errors the shortfall is even larger (−7.6 pp). In addition, deflationary periods in the UK are associated with higher unemployment, as well as lower growth in industrial production, consumption, and investment.

Why do these results differ from the existing literature? To shed some light on this question, I analyze GDP per capita growth for eleven countries. The varying quality of the price data among these countries may introduce varying degrees of misclassification and deflator biases. There is a positive correlation between the volatility of GDP growth and the volatility of inflation. In addition, there is a positive link between the volatility of real GDP growth and the average growth rate during periods of deflation. These patterns provide suggestive evidence that the deflator bias affects cross-country studies, because real GDP may be more volatile for those countries with more serious measurement errors in prices.

Because prices frequently declined under metallic standards, studies on the link between real activity and deflation often use 19th century data which suffer from more serious measurement errors than modern statistics (see, e.g., Romer, 1986a). Whether deflation is associated with lower real activity is controversial. Atkeson and Kehoe (2004), Bordo and Filardo (2005), and Borio et al. (2015) find, over a large number of countries, a weak link between real activity and deflation. Eichengreen et al. (2016) report that the link becomes stronger when using wholesale prices instead of consumer prices. One explanation for the weak link is that many 19th century deflationary episodes were driven by advances in productivity (see Friedman & Schwartz, 1963; Beckworth, 2007). Another explanation emphasizes that the economy adapted more easily to adverse demand shocks because labor and product markets were more flexible (see Bayoumi & Eichengreen, 1996, and references therein). This paper argues in favor of a third possibility: measurement errors in price data bias estimates of economic growth during deflationary periods.

Most researchers readily acknowledge measurement errors in 19th century price data as a caveat in their empirical work (see, e.g., Barsky, 1987; Benati, 2008). Some studies therefore prefer relatively accurate wholesale price indices (see, e.g., Barsky, 1987; Cogley & Sargent, 2015; Eichengreen et al., 2016).[3] Others examine the robustness of the results using alternative price indices (see Margo, 2000, Chapter 2). I argue that all 19th century price indices are imperfect measures. However, we can jointly exploit these indices to improve estimates of economic growth during deflation.

The paper is also related to Romer (1986a, 1986b), Allen (1992), and Hanes (1998). They apply 19th century methodologies to modern data to assess the properties of mismeasured historical statistics. Based on these modern replications, Cogley and Sargent (2015) exploit the overlap between modern data and modern replication to control for persistent measurement errors in historical estimates of trend inflation. Similarly, I estimate the misclassification rate for modern data to control for measurement errors in the historical analysis.[4]

There is an extensive literature showing that measurement errors account for missing growth in modern data. Modern price data overestimate inflation for several reasons, and as a result, we underestimate real economic growth. The Boskin Commission (1996) shows that missing quality adjustments in modern CPI data cause inflation to be overestimated by 1 pp. Goolsbee and Klenow (2018) find that the bias is larger for newly emerging online products. Aghion et al. (2019) suggest that disappearing products introduce upward bias in inflation and, therefore, downward bias in economic growth. Similarly, this paper argues that measurement errors account for the missing shortfall of economic growth during deflation.

The remainder of the paper is structured as follows. I first describe the properties of the misclassification and deflator biases and propose remedies. Then, I discuss the main deficiencies of 19th century price data, as well as the misclassification rates using modern replications. Thereafter, I assess the shortfall in real activity during deflation and assess the size of the bias due to measurement errors. The last section concludes the paper.

[3] Consumer price indices for the 19th century are regarded as less accurate because retail price data are particularly scarce (see Hoover, 1958).

[4] This strategy is inspired by a microeconometric literature using external information from validation surveys to assess and control for measurement errors. For example, Card (1996) uses external information from a 1977 validation study to set the misclassification rate in reported union status in an analysis for the period 1987–1988. Additionally, Bound and Krueger (1991) use an earnings survey matched with administrative data to examine the severity and properties of reporting error.

2 | MISCLASSIFICATION AND DEFLATOR BIASES

Measurement errors in historical data are often believed to attenuate the correlation between macroeconomic time series.[5] This intuition fails in two ways when assessing the relationship between real activity and deflation. First, because the mismeasured independent variable is a binary indicator, the measurement errors are necessarily correlated with the true variable. Second, if the GDP deflator suffers from similar deficiencies as the price index underlying the binary indicator, the measurement errors of the dependent and independent variables are negatively correlated.

How large was the average shortfall in real activity when prices were falling? How robust was real activity when prices were rising? I cast these questions in a linear regression model:[6]

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

where y_t represents real economic growth, $d_t \equiv \mathbf{1}_{\{\pi_t<0\}}$ is a binary deflation indicator, and ε_t is an i.i.d. error term. A negative β implies that deflation is associated with a decline in real activity. A positive α implies that real activity is buoyant when prices are rising. $\alpha + \beta$ measures average real activity when prices are falling.

If inflation and real activity are measured with error we estimate:

$$\begin{aligned} \tilde{y}_t &= \alpha + \beta x_t + \epsilon_t \\ \epsilon_t &= \varepsilon_t - \beta(x_t - d_t) + (\tilde{y}_t - y_t), \end{aligned} \tag{2}$$

where $\tilde{y}_t$ is an error-ridden measure of real activity and $x_t \equiv \mathbf{1}_{\{\tilde{\pi}_t<0\}}$ is an erroneous binary indicator based on mismeasured inflation ($\tilde{\pi}_t$).

The error term (ϵ_t) includes the measurement errors of real activity ($\tilde{y}_t - y_t$) and the classification errors ($x_t - d_t$). The classification errors are necessarily correlated with the binary indicator d_t (see Aigner, 1973).[7] In addition, the errors in real activity are negatively correlated with the erroneous binary indicator if the deflator suffers from similar deficiencies as the price index underlying the binary indicator. Therefore, we cannot consistently estimate Equation 2 by OLS.

2.1 | Properties of the biases

To characterize the direction and size of the biases, I derive the probability limit of the OLS estimator. Assuming that the measurement errors in real activity are unrelated to the measurement errors in inflation we obtain (see Aigner, 1973):

$$\begin{aligned} \operatorname{plim} \widehat{\alpha}_{ols} &= E[\tilde{y}_t|x_t = 0] \\ &= \alpha + \beta P[d_t = 1|x_t = 0] \\ \operatorname{plim} \widehat{\beta}_{ols} &= E[\tilde{y}_t|x_t = 1] - E[\tilde{y}_t|x_t = 0] \\ &= \beta(1 - P[d_t = 0|x_t = 1] - P[d_t = 1|x_t = 0]) \ . \end{aligned} \tag{3}$$

If $\beta < 0$ we underestimate real activity during inflationary periods ($\operatorname{plim} \widehat{\alpha}_{ols} < \alpha$). In addition, we overestimate the shortfall in real activity during deflation ($\operatorname{plim} \widehat{\beta}_{ols} > \beta$). In other words, deflationary periods appear more benign, and inflationary periods appear less buoyant.

The bias becomes larger if the measurement errors in real activity are negatively correlated with the measurement errors in inflation. Suppose that we measure nominal GDP growth without error (n_t).[8] However, we use an error-ridden price index as a deflator. Measured real GDP growth then approximately equals $\tilde{y}_t \approx n_t - \tilde{\pi}_t$. Therefore, the measurement errors in real GDP growth are negatively correlated with the measurement errors in inflation: $\tilde{y}_t - y_t = n_t - \tilde{\pi}_t - (n_t - \pi_t) = -(\tilde{\pi}_t - \pi_t)$.

[5] The intuition follows from the well-known attenuation bias: if the independent variable is measured with i.i.d. errors, the OLS estimate of the coefficient is biased towards zero (see Griliches, 1986; Hausman, 2001). For example, Jordà et al. (2016) state the following: “We also confirm that consumption and investment are procyclical with output. This comovement seems to increase over time, potentially reflecting better measurement.”
[6] See Reinhart and Rogoff (2010), Borio et al. (2015), Eichengreen et al. (2016), and Jordà et al. (2016) for historical studies reporting descriptive statistics on various economic relationships that can be cast in this model.
[7] Pakes (1982) shows that Wald-type estimators are generally biased if the underlying data are subject to error. In such estimators, we divide observations into groups with above- and below-median observations on the independent variable and then fit a line through the group means.
[8] Allowing for i.i.d. measurement errors in nominal GDP does not change the probability limits.

The probability limits of the OLS estimator read:

$$\begin{aligned}\text{plim } \hat{\alpha}_{ols} &= \alpha + \beta P[d_t = 1|x_t = 0] - E[(\tilde{\pi}_t - \pi_t)|x_t = 0] \\ \text{plim } \hat{\beta}_{ols} &= \beta(1 - P[d_t = 0|x_t = 1] - P[d_t = 1|x_t = 0]) \\ &\quad + E[(\tilde{\pi}_t - \pi_t)|x_t = 0)] - E[(\tilde{\pi}_t - \pi_t)|x_t = 1] \ .\end{aligned} \tag{4}$$

The bias increases because $E[(\tilde{\pi}_t - \pi_t)|x_t = 0] > 0 > E[(\tilde{\pi}_t - \pi_t)|x_t = 1]$. Intuitively, some inflationary (deflationary) periods are caused by positive (negative) measurement errors in inflation $(\tilde{\pi}_t - \pi_t)$. These measurement errors lead to lower (higher) real GDP growth because of the error-ridden deflator.

To assess the properties and severity of the biases I conduct a simulation exercise, assuming that the error-ridden inflation measure depends linearly on actual inflation, distorted by an i.i.d. error (ω_t):

$$\tilde{\pi}_t = \rho_0 + \rho_1 \pi_t + \omega_t \ , \tag{5}$$

The functional form allows for a mismeasured intercept (ρ_0) and slope (ρ_1). In addition, the binary indicator depends on a threshold c $(x_t \equiv \mathbf{1}_{\{\tilde{\pi}_t < c\}})$.[9] In the baseline, I examine the impact of i.i.d. measurement errors (ρ_0=0, ρ_1=1, c=0). I set the overall volatility of observed inflation to $\sigma = \sqrt{\sigma_\pi^2 + \sigma_\omega^2} = 6$, while varying the signal-to-noise ratio $(\sigma_\pi^2/\sigma_\omega^2)$ from 0 to 6.[10] Then, I vary the threshold $(c = 5)$, the intercept $(\rho_0 = 5)$, and the slope parameter $(\rho_1 = 3)$. In addition, I examine the impact of correlated measurement errors in the dependent variable relative to the baseline by setting $\tilde{y}_t - y_t = -(\tilde{\pi}_t - \pi_t) = -\omega_t$.

Figure 1, panels (a) and (b) show the simulated probability limits, assuming that measurement errors in real activity are unrelated to measurement errors in inflation. Panels (c) and (d) show the probability limits with correlated measurement errors. The solid horizontal lines depict the true values of the parameters.[11]

Five observations emerge. First, the misclassification bias drives the probability limits of the coefficients in opposite directions (panels a and b).[12] However, the bias in β is larger in absolute size. Second, the bias becomes larger when increasing the threshold c (solid line). Choosing a threshold that differs from the true unconditional average implies that we classify a rare event. It therefore becomes more likely that the indicator misclassifies the event.[13] Third, a mismeasured intercept exacerbates the bias. If we systematically over- or underestimate inflation we will always misclassify some episodes. Therefore, this bias does not vanish completely even if we let the signal-to-noise ratio go to infinity. Fourth, increasing the slope coefficient reduces the misclassification bias. Fifth, if we allow for correlated measurement errors between the independent and dependent variables, the bias becomes larger (panels c and d). The probability limits are of the wrong sign, except for a very high signal-to-noise ratio.

These results have implications for studies analyzing the link between deflation and real activity. For countries with measurement errors in price data, we underestimate (overestimate) real activity during inflationary (deflationary) episodes. Moreover, measurement errors in the mean bias the OLS estimates.[14] This bias is likely relevant because, even at present, we overestimate inflation because of technological progress, creative destruction, and product substitution (Aghion et al., 2019; Boskin Commission, 1996; Goolsbee & Klenow, 2018). Then, the bias is more pronounced if we examine rare events, for example, particularly severe deflationary episodes. The bias is mitigated, however, if we obtain an inflation measure that places greater weight on volatile, but well-measured, prices. This may explain the contrasting results obtained by Borio et al. (2015) using consumer prices and Eichengreen et al. (2016) using wholesale prices. Finally, the deflator bias likely changes the sign of the relationship between GDP growth and deflation. Therefore, the intuition that measurement errors merely attenuate the estimates does not hold.

[9] For 19th century inflation, the natural threshold is zero. However, some studies report 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%.
[10] The overall volatility corresponds to inflation volatility in the 19th century US.
[11] Online Appendix C, Figure C.1, shows simulations for average GDP growth during deflation $(\alpha + \beta)$.
[12] The misclassification bias in the baseline is larger than the attenuation bias with a continuous independent variable for signal-to-noise ratios larger than unity. Forming a binary indicator therefore exacerbates the bias if the measurement errors are more volatile than actual inflation. Data transformations, for example squaring a continuous regressor, often exacerbate the bias from measurement errors (see Griliches, 1986). Kreider (2010) shows that with arbitrary forms of classification error moderate rates of misclassification can lead to even more serious biases.
[13] The same phenomenon occurs in diagnosing rare illnesses. A small rate of false positives may imply that most of the individuals who test positive are in fact healthy.
[14] This contrasts with the continuous case where the estimate of the slope coefficient is not affected by a mismeasured intercept (see Griliches, 1986).

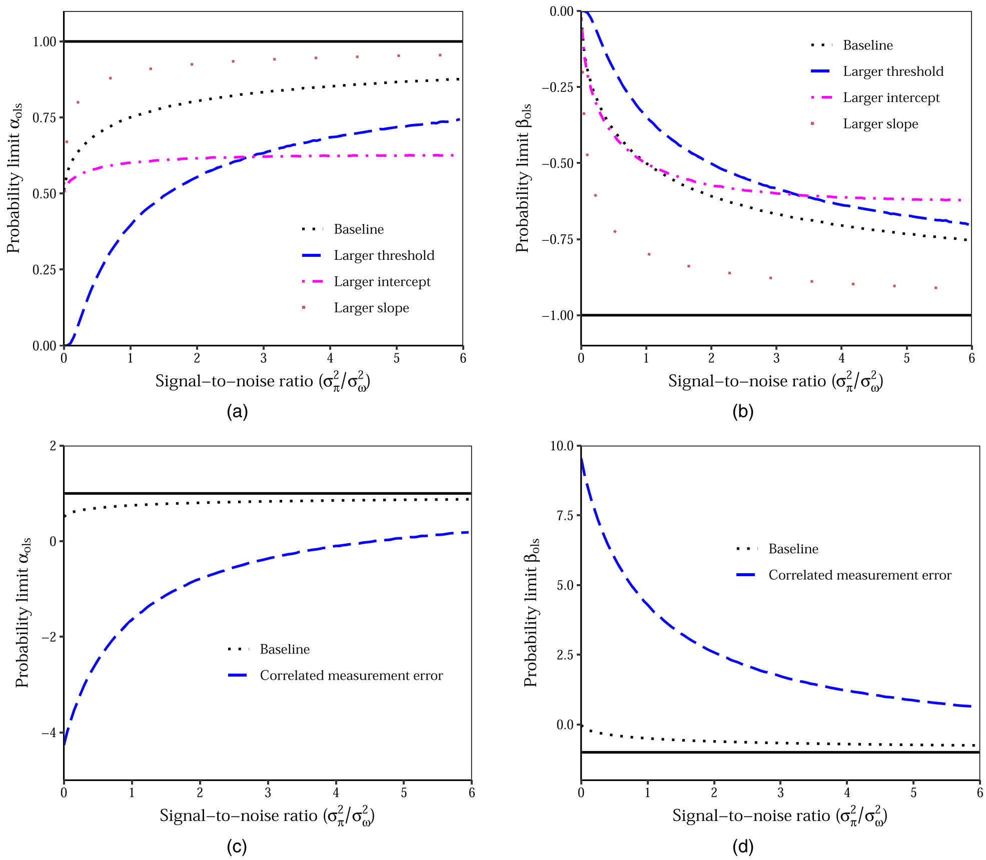

FIGURE 1 Simulated probability limit. The figure shows the probability limit of the OLS estimator of $y_t = \alpha + \beta x_t + \epsilon_t$, where $x_t \equiv \mathbf{1}_{\{\tilde{\pi}_t < c\}}$, as a function of the signal-to-noise ratio $(\sigma^2_\pi/\sigma^2_\omega)$. The solid horizontal lines give the true values of $\alpha = 1$ and $\beta = -1$. The error-ridden inflation rate depends linearly on the well-measured inflation rate ($\tilde{\pi}_t = \rho_0 + \rho_1 \pi_t + \omega_t$). The well-measured inflation rate (π_t) and the measurement errors (ω_t) are assumed to be identically and independently normally distributed with zero mean. The baseline simulation assumes that $c = 0, \rho_0 = 0, \rho_1 = 1$, and $\sigma = \sqrt{\sigma^2_\pi + \sigma^2_\omega} = 6$. The other simulations assume a larger threshold ($c = 5$), a larger intercept ($\rho_0 = 5$), and a larger slope ($\rho_1 = 3$). Panels (a) and (b) show the probability limit with uncorrelated measurement errors in the dependent variable (see Equation 3). Panels (c) and (d) show the probability limit with negatively correlated measurement errors in the dependent variable (see Equation 4) [Colour figure can be viewed at wileyonlinelibrary.com]

2.2 | Bounds and consistent estimates

To address the deflator bias, the empirical analysis focuses on real activity measures, which are directly estimated in real terms, and on countries with arguably well-measured data. Assuming that the remaining measurement errors in real activity and inflation are unrelated we can resolve the misclassification bias using a second error-ridden binary indicator ($z_t \equiv \mathbf{1}_{\{\hat{\pi}_t < 0\}}$).[15]

2.2.1 | Conditional independence

Kane et al. (1999) and Black et al. (2000) derive bounds for the true coefficient, assuming that the misclassification errors are independent conditional on the true deflation indicator. Black et al. (2000) show that the OLS estimate of β_{11} in:

$$\tilde{y}_t = \alpha + \beta_{11}\mathbf{1}_{\{x_t=1,z_t=1\}} + \beta_{10}\mathbf{1}_{\{x_t=1,z_t=0\}} + \beta_{01}\mathbf{1}_{\{x_t=0,z_t=1\}} + \epsilon_t \quad , \tag{6}$$

is closer to the true value of β. If we obtain two independent signals that the price index declines, the probability of an erroneous classification is lower. The OLS estimate is still biased, however, because both indicators may simultaneously misclassify a period. Therefore, this approach yields an upper bound if $\beta<0$.

[15] See Online Appendix A for a detailed discussion.

We can estimate another bound using the second binary indicator as an instrumental variable.[16] The probability limit of the IV estimator equals (see Kane et al., 1999):

$$\text{plim } \hat{\beta}_{IV} = \frac{\beta}{1 - P[x_t = 0|d_t = 1] - P[x_t = 1|d_t = 0]} . \tag{7}$$

Therefore, IV yields a lower bound if $\beta < 0$.

Finally, we can consistently estimate α and β using GMM (see and Online Appendix A Black et al., 2000; Kane et al., 1999). Note that we can estimate seven empirical moments from the data: three sampling fractions of the binary indicators, and the conditional means of the dependent variable for each of the four combinations of the binary indicators. From these empirical moments, we have to estimate seven parameters: two model coefficients, four misclassification rates, and the actual rate of deflation. Therefore, the model is just identified. Online Appendix A shows how to recover an estimate of the bias as a nonlinear function of the underlying GMM estimates.

2.2.2 | Conditional dependence

In the present application the conditional independence assumption is likely too strong.[17] Many historical price indices suffer from the same deficiencies. I therefore propose a second identification strategy.

Without the conditional independence assumption, we have to identify nine parameters from seven empirical moments. Therefore, we have to impose two additional restrictions. I assume that the joint misclassification probabilities ($P[x_t = 0, z_t = 0|d_t = 1]$, $P[x_t = 1, z_t = 1|d_t = 0]$) are known.[18] If the measurement errors in the binary indicators are only weakly related, these probabilities are small. The next section assesses how small these probabilities are using modern replications of 19th century price indices.

3 | DATA

This section presents the properties of two US price indices, gauges the joint misclassification probabilities using modern replications, and discusses the real activity data used in the main analysis.

3.1 | A composite CPI for the USA

The composite CPI of Officer and Williamson (2016) likely represents the most accurate US CPI at a given point in time. Nevertheless, the retrospectively constructed segments suffer from various methodological deficiencies.

Before 1800, David and Solar (1977) use wholesale prices to approximate prices at the retail stage.[19] In addition, these prices stem from a limited geographical area, namely Philadelphia. After 1800, the price index is based on retail prices from Adams (1939). Limited geographical coverage remains an issue because these prices were recorded by farmers in Vermont. After 1850, Hoover (1960) uses data from the Weeks (1886) Report, a retrospective survey covering a wider geographical area and a broad range of retail products. Even the comprehensive Weeks Report, however, suffers from a relatively small number of individual price quotes.[20] Hoover (1958) emphasizes that retail price data are particularly scarce for the period 1880–1890, after the Weeks Report ends and before the US Bureau of Labor Statistics began to collect retail prices for food items. Therefore, Long (1960) approximates the prices of several items, rent for example, by a linear interpolation over the entire 1880s. Information on rent for housing is generally scarce. For 1860–1880 Lebergott (1964) constructs a reproduction cost index by equally weighing the cost of construction materials and wages for low-skilled workers. Finally, a general defect is the lack of service prices. For example, the indices developed by Lebergott (1964) and Hoover (1960) comprise only a few service items: rent, shoe repairs, and physician fees.[21]

[16] Kane et al. (1999) emphasize that IV does not resolve the bias because the misclassification error is necessarily correlated with the true regressor.
[17] The conditional independence assumption is often used in applications where the education level is reported in different surveys by the employer and the employee (see Black et al., 2000). Therefore, it is reasonable to assume that the respondents make independent reporting errors.
[18] I would like to thank Bo Honoré for bringing this possibility to my attention.
[19] See Table B.1 Online Appendix B.
[20] See Online Appendix B.2 for a detailed discussion.
[21] Similar issues plague CPI estimates well into the 20th century. The US CPI began to be published on a monthly basis in 1940; even then, many service prices were still collected quarterly. Previously, the CPI was only published at irregular intervals or even only for December. During the period 1913–1921, the BLS retrospectively estimated a monthly CPI, interpolating prices for many items that were not collected monthly (see Officer, 2014).

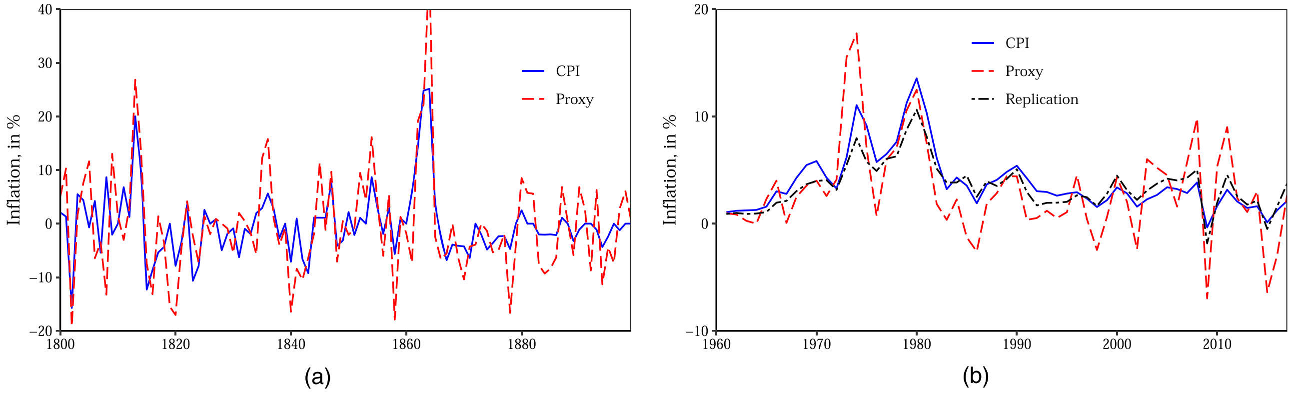

FIGURE 2 CPI inflation, proxy, and modern replications. The proxy combines wholesale and producer price indices with consumer expenditure weights from Hoover (1960) for the 19th century, as well as CPI weights for 2017 from the BLS for the post-WWII period. CPI inflation stems from Officer and Williamson (2016). The replication uses modern data to reconstruct the 19th-century CPI constructed by Hoover (1960). [Colour figure can be viewed at wileyonlinelibrary.com]

3.2 | A new proxy for the USA

To obtain a second binary indicator I construct a novel proxy based on wholesale prices from Warren and Pearson (1933) and Hanes (1998). The proxy combines price indices for five commodity groups weighed by constant consumer expenditure weights.[22] Using wholesale prices has three advantages: The data stem from distinct sources, and therefore, the measurement errors are to some extent independent from those in the composite CPI;[23] the underlying data are accurate because the information stems from newspapers and company archives (see Hoover, 1958); finally, the misclassification bias may be less severe because wholesale prices fluctuate more strongly than consumer prices.

Panel (a) in Figure 2 shows 19th century inflation measured by the composite CPI and the proxy. The means of the two series are identical (−0.2%). In addition, the series display similar turning points, and the correlation is sizable (0.74). However, the standard deviation is higher for the proxy than for the composite CPI (9.9 compared to 6.2). Despite their high correlation, the two indicators yield a different classification of inflation or deflation for 29% of the observations.

3.3 | Misclassification rates from modern replications

Despite the use of distinct data sources, the measurement errors affecting the CPI and the proxy are likely to be correlated because the two indices suffer from similar deficiencies. Thus, the conditional independence assumption is violated. Although we cannot directly test this assumption, we can replicate the proxy and the retrospective estimate of CPI inflation using modern data. Because we observe both, the erroneous and well-measured series, we can gauge the joint misclassification rates.[24]

I replicate the 19th century CPI developed by Hoover (1960) using modern data for the period 1960–2017.[25] In addition, I construct the proxy for the post-WWII era using modern PPI data from the US Bureau of Labor Statistics. Panel (b) in Figure 2 shows that the proxy is correlated with post-WWII CPI inflation and reflects major up- and downturns. Mirroring the 19th century data, the growth rate of the proxy is more volatile than that of the actual CPI and the replication.[26]

[22] See Online Appendix B.4. The proxy covers more than 70% of a 19th century consumption basket. The most important missing item is rent (18% of the consumption basket).
[23] Ideally, the errors of the CPI and the proxy would be independent. A necessary condition is that the underlying data sources differ. Online Appendix B.4 shows that this is indeed the case for most of the 19th century.
[24] That the modern CPI is measured without error is not exactly true. The Boskin Commission (1996) shows that the CPI may underestimate actual CPI inflation because of neglected changes in quality and product substitution (see Goolsbee & Klenow, 2018, for a more recent study). However, the sampling error in modern CPIs is small (see Shoemaker, 2014).
[25] Online Appendix B.5 provides a detailed description. In addition, Online Appendix B.6 shows that the signal-to-noise ratios lie between 1 and 6 for the modern replications. Based on these modern replications and the simulation exercise in Figure 1, we therefore expect to find relevant misclassification bias.
[26] The modern replication, however, understates the importance of measurement errors for two reasons. The Hoover (1960) consumption basket covers a wide range of expenditure items. Other segments are likely affected by more serious measurement errors because they cover fewer items. Moreover, the post-WWII data are based on comprehensive surveys and are therefore more accurately measured than their historical counterparts. For example, the number of observations in modern price data is larger and therefore sampling error substantially smaller (see Online Appendix B.2 for a discussion).

TABLE 1 Misclassification rates

	Replication		Proxy		Product		Joint	
	$\boldsymbol{d} = 1$	$\boldsymbol{d} = 0$	$\boldsymbol{d} = 1$	$\boldsymbol{d} = 0$	$\boldsymbol{d} = 1$	$\boldsymbol{d} = 0$	$\boldsymbol{d} = 1$	$\boldsymbol{d} = 0$
1960-2017 zero threshold	0.00	0.02	0.00	0.12	0.00	0.00	0.00	0.02
1960-2017 higher threshold	0.22	0.20	0.27	0.40	0.06	0.08	0.11	0.20

Note: Misclassification rates based on actual CPI inflation, the replication, and the proxy. Each panel comprises rates of misclassifying deflationary ($P[x_t = 0|d_t = 1]$) and inflationary ($P[x_t = 1|d_t = 0]$) periods. The third panel reports the product of the misclassification rates for the replication and the proxy, which would correspond to the joint misclassification rate if the two indicators were conditionally independent (i.e. $P[x_t = 1, z_t = 1|d_t = 0] = P[x_t = 1|d_t = 0]P[z_t = 1|d_t = 0]$). The last panel presents the joint misclassification rates. The second and third rows vary the threshold (average inflation instead of 0) and the sample period.

Table 1 provides misclassification rates for the replication and the proxy to gauge whether conditional independence is a reasonable assumption and how likely the proxy and the CPI replication are to misclassify an inflationary or deflationary episode. Because the inflation mean affects misclassification rates and because the inflation mean was substantially lower during the 19th century, the table also presents results for a higher threshold set to the unconditional mean of CPI inflation.

For the entire sample, the replication never misclassifies deflationary periods and rarely misclassifies inflationary periods. The proxy is less accurate: it misclassifies 12% of inflationary periods. The third panel in Table 1 reports the product of the individual misclassification rates, which should be equal to the joint misclassification rates if the indicators are conditionally independent. For the entire sample, the product and the joint misclassification rates are close to zero.

These results are not surprising because inflation was relatively high during this sample period and the replication is less volatile than actual CPI inflation. Therefore, the second row reports misclassification rates with a threshold set to the unconditional mean of well-measured inflation. The individual misclassification rates rise to between 20% and 40% for the proxy and the replication. By contrast, the joint misclassification rates amount to only 11% (deflation) and 20% (inflation). This confirms that exploiting the information from two error-ridden indicators mitigates the misclassification bias. Under the conditional independence assumption, however, these rates should be even lower at 6% (deflation) and 8% (inflation).

3.4 | Real activity data

The choice of real activity data matters because of the deflator bias. I therefore proceed in three steps. First, I use US industrial production from Davis (2004). His series is a quantity-based measure directly estimated in real terms and is therefore not affected by measurement errors in price indices.[27] Second, I use various real activity measures for the UK because these data are arguably of better quality than those for other countries.[28] Third, I extend the analysis using GDP growth for eleven countries. The severity of measurement errors in prices, as well as the correlation with errors in the GDP deflator, may vary across countries. I therefore examine whether the cross-country variation in the estimates is consistent with differences in the misclassification and deflator biases.

4 | THE COST OF DEFLATION REVISITED

4.1 | Evidence for the 19th century USA and UK

Table 2 provides estimates of average industrial production growth during periods with rising prices (α), as well as the average shortfall and the average growth rate during periods with falling prices (β, $\alpha + \beta$). The first column uses the CPI developed by Officer and Williamson (2016). According to this classification, industrial production growth amounts to 6.5% during inflationary periods. During periods of deflation, industrial production growth is not statistically significantly lower. Indeed, growth under deflation amounts to a robust 4.5%. Because the CPI is likely measured with error, these estimates provide an upper bound for the negative association between economic growth and deflation.

The second to fourth columns provide bounds and point estimates under the assumption that the binary indicators based on the composite CPI and the proxy are conditionally independent. The shortfall in industrial production growth

[27] Even US real GDP may not be strongly affected by the deflator bias. As Johnston and Williamson (2019) note, actual estimates are available only for decennial benchmark years. Between these years, the annual real GDP figures rely heavily on the industrial production index developed by Davis (2004). Therefore, the deflator bias may mostly affects the benchmark years, not the other periods.

[28] The data sources are described in Online Appendix B.7.

TABLE 2 USA industrial production growth during periods of inflation and deflation (1800–1899)

	Baseline	Cond. independence			Cond. dependence
Model parameters:					
$\alpha = E[y\|\pi > 0]$	6.51***	7.80***	7.79***	10.25***	9.72***
	(1.02)	(1.07)	(1.28)	(1.96)	(1.87)
$\beta = E[y\|\pi < 0] - E[y\|\pi > 0]$	−2.06	−4.35***	−4.47***	−9.68***	−7.60***
	(1.49)	(1.38)	(1.64)	(3.56)	(2.35)
$\alpha + \beta = E[y\|\pi < 0]$	4.45***	3.45***	3.32***	0.56	2.12*
	(1.01)	(0.89)	(1.07)	(1.90)	(1.23)
$P[\pi < 0]$			0.51***		0.56***
			(0.17)		(0.14)
Bias estimates:					
plim $\hat{\alpha} - \alpha$			−1.28		−2.84*
			(0.94)		(1.58)
plim $\hat{\beta} - \beta$			1.41**		3.15**
			(0.64)		(1.49)
plim $\hat{\alpha} + \hat{\beta} - \alpha - \beta$			1.13		2.33**
			(0.75)		(1.00)
N	100	100	100	100	100
Bound	Upper	Upper	Point	Lower	Point
Method	OLS	OLS	GMM	IV	GMM
Indicator	CPI	CPI, proxy	CPI, proxy	CPI, proxy	CPI, proxy

Note. Model: $y_t = \alpha + \beta x_t + \epsilon_t$ where $x_t \equiv \mathbf{1}_{\{\tilde{\pi}_t<0\}}$. Model parameters: Mean growth rate under inflation α; shortfall under deflation β; mean growth rate under deflation $\alpha+\beta$; probability of deflation ($P\pi < 0$). Bias estimates: Bias if we were to only use the CPI; calculated based on the underlying GMM estimates, with standard errors computed using the delta method. Baseline: OLS estimates using the CPI. Conditional independence: Bounds and point estimates using the CPI and proxy, assuming conditional independence. Conditional dependence: Point estimates using the CPI and proxy, assuming that the joint misclassification probabilities equal 0.15. HAC-robust standard errors are given in parentheses. ***, **, * denotes significance at the 1%, 5%, 10% level.

during periods of deflation becomes more pronounced. We find a statistically significant shortfall of −4.5 pp. Moreover, the bounds suggest that the actual coefficient lies between −4.4 pp and −9.7 pp. The last column relaxes the conditional independence assumption assuming that the joint misclassification rates equal 15%. The shortfall under deflation becomes even larger (−7.6 pp).

It is worth emphasizing that the classification error does not bias the unconditional mean of industrial production growth. It distorts the allocation of growth to periods with rising and falling prices, as well as the share of deflationary periods. Therefore, it is possible that deflationary periods were, on average, still associated with robust growth. The mean growth rate during periods of deflation declines when controlling for measurement errors. In no specification, however, is the growth rate significantly negative.

Are the biases statistically significant? Although we cannot directly compare the standard errors across models, GMM allows us to estimate the bias as a nonlinear function of the underlying coefficients.[29] The lower panel shows the bias estimates for the CPI. Under the conditional independence assumption, the shortfall in real activity is significantly biased by 1.4 pp. Allowing for conditional dependence we find significant biases in the growth rate under inflation (−2.8 pp, 10% level) and the growth rate under deflation (2.3 pp, 5% level). Therefore, accounting for measurement errors yields a statistically and economically different assessment of the relationship between real activity and deflation.

Analyzing the USA has the drawback that the country was little developed during the 19th century. Data are therefore necessarily scarce and the economy tilted towards agriculture. I therefore examine the UK, which offers the advantage that various well-measured real activity measures exist. I use a CPI for the baseline classification and a wholesale price index as a proxy.[30] Table 3 shows the change in various real activity measures during deflationary episodes. During deflationary periods the unemployment rate increases, while growth in consumption, investment, and industrial production declines. In all specifications, the estimates become larger in absolute size when exploiting a second deflation indicator. Therefore, even for relatively well-measured UK data we find a relevant misclassification bias.

[29] See Online Appendix A for a derivation.

[30] A detailed examination of the quality of the data, along the lines of the previous exercise for the USA, is beyond the scope of this paper. I therefore assume the same misclassification rates as for the USA

TABLE 3 Change in UK economic activity during periods of deflation (1830–1899)

	Baseline	Cond. independence			Cond. dependence
Industrial production growth	−1.42	−2.82**	−3.36**	−5.78*	−5.30**
	(1.06)	(1.36)	(1.54)	(2.98)	(2.67)
Unemployment rate	2.03***	2.27***	2.72***	3.18**	3.99***
	(0.55)	(0.62)	(0.81)	(1.32)	(1.12)
Consumption growth	−1.05**	−1.40**	−1.55**	−2.24*	−2.55**
	(0.45)	(0.57)	(0.69)	(1.21)	(1.15)
Investment growth	−7.15**	−10.58**	−11.95**	−18.45*	−19.41**
	(3.62)	(4.74)	(5.56)	(9.64)	(9.17)
N	70	70	70	70	70
Bound	Upper	Upper	Point	Lower	Point
Method	OLS	OLS	GMM	IV	GMM
Indicator	CPI	CPI, proxy	CPI, proxy	CPI, proxy	CPI, proxy

Note. Model: $y_t = \alpha + \beta x_t + \epsilon_t$ where $x_t \equiv \mathbf{1}_{\{\tilde{\pi}_t<0\}}$. Model parameters: Change in economic activity under deflation β. Baseline: OLS estimates using the CPI. Conditional independence: Bounds and point estimates using the CPI and proxy, assuming conditional independence. Conditional dependence: Point estimates using the CPI and proxy, assuming that the joint misclassification probabilities equal 0.15. HAC-robust standard errors are given in parentheses. ***, **, * denotes significance at the 1%, 5%, 10% level.

4.2 | Cross-country evidence

Why do these results differ from existing studies? First, I analyze a different sample period. However, the results remain qualitatively similar when focusing on various subsamples.[31] Second, other studies examine the link between inflation and real activity using GDP growth. However, I also find a significant decline in real US GDP growth.[32] Finally, I focus on two countries rather than on a large cross-country data set. If GDP data for other countries are more affected by the deflator bias, a weaker, or even positive, link between deflation and real GDP growth may emerge. In what follows, I provide cross-country evidence in line with this interpretation. It is worth pointing out that the correlations discussed in this section provide only suggestive evidence of the misclassification and deflator biases. The cross-country patterns may also stem from differences in, for example, supply shocks simultaneously affecting real GDP and prices. It is beyond the scope of the paper, however, to judge the quality of the underlying data for all these countries.

Figure 3 panel (a) shows OLS estimates using GDP per capita growth for eleven countries.[33] GDP growth is higher during deflationary periods for Portugal, Finland, Switzerland, and Spain. However, there is a negative association for the USA, UK, and Sweden. For countries with a high β, α tends to be low. Recall that measurement errors in prices bias the probability limits of the two coefficients in opposite directions (see Figure 1). One interpretation of the cross-country pattern therefore is that the misclassification and deflator biases are more serious for Portugal and Switzerland than for the UK or the USA.[34]

Panel (b) shows the relationship between the volatility of inflation and real GDP growth. The variance is computed relative to the variance in the UK, assuming that UK inflation and GDP growth are relatively well measured. If the same measurement errors appear in the CPI and the deflator, we expect the volatility of CPI inflation and real GDP growth to be positively correlated across countries. Indeed, there is a positive relationship between the two.

In addition, we expect countries with more volatile measurement errors to exhibit the largest biases. Panels (c) and (d) show that there is a significant relationship between the shortfall under deflation (β) and the variance of real GDP growth. However, the correlation is only significant at the 10% level for the variance of CPI inflation. When excluding Sweden, which appears to be an outlier, the relationship is statistically significant.[35]

[31] Although the coefficients are less precisely estimated before 1870. See Online Appendix C.
[32] See Online Appendix C. This does not come as a surprise because annual US GDP estimates rely heavily on the industrial production data constructed by Davis (2004).
[33] I am very grateful to an anonymous referee for providing these data.
[34] At least for Switzerland the 19th century composite CPI is only a rough estimate (see David & Ritzmann, 2012). During various episodes the index uses wholesale prices, covers a limited geographical area, and is very noisy (see Kaufmann, 2019; Studer & Schuppli, 2008)
[35] See Online Appendix C.

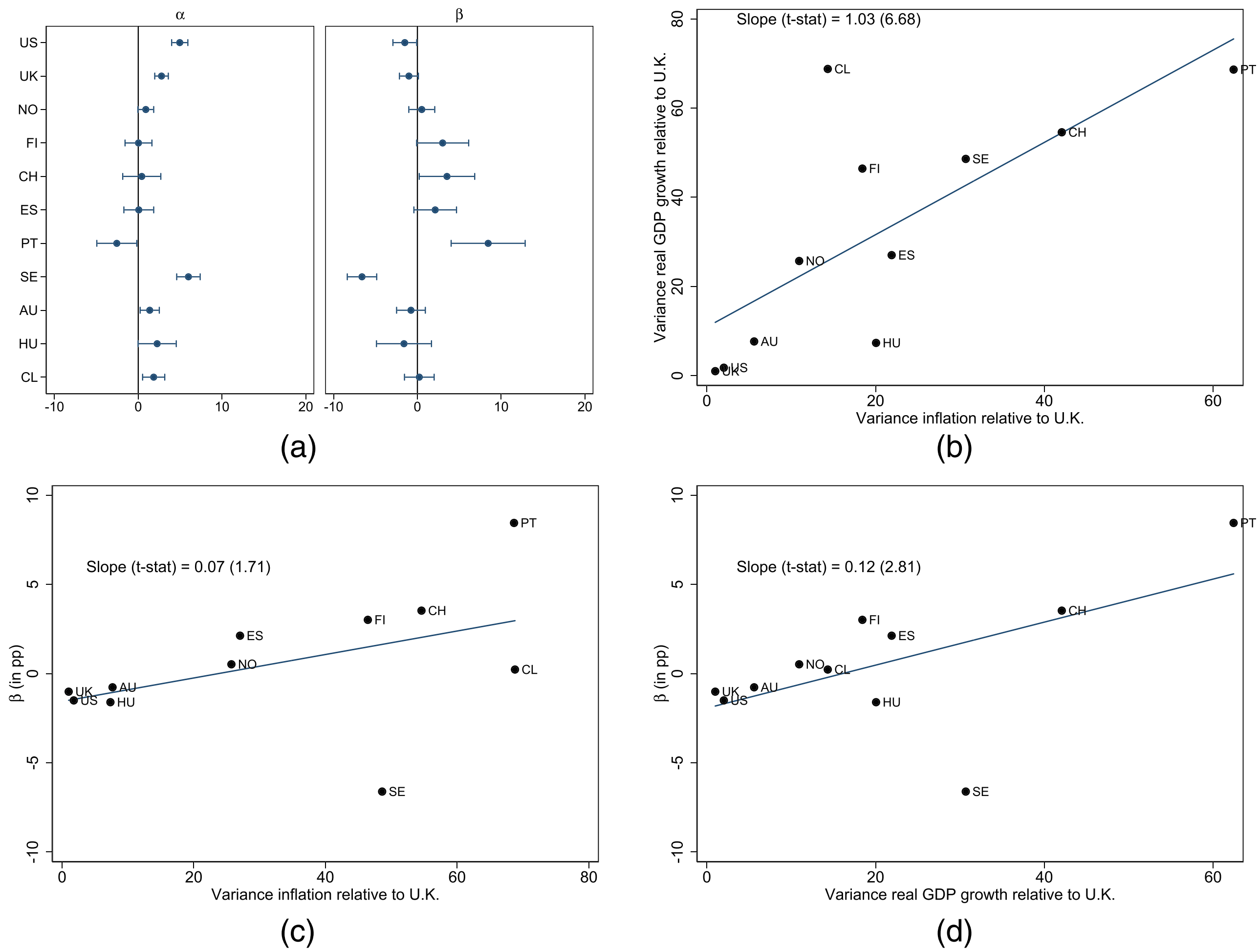

FIGURE 3 Volatility of historical data and OLS estimates. Model: $y_t = \alpha + \beta x_t + \epsilon_t$ where $x_t \equiv \mathbf{1}_{\{\tilde{\pi}_t<0\}}$. (a) Real GDP growth under inflation (α) and shortfall under deflation (β), with HAC-robust 95% confidence intervals. (b) Relationship between the volatility of inflation and GDP growth. (c,d) Relationship between the volatility of macroeconomic data and the OLS estimate of shortfall under deflation (β). All volatilities normalized by the volatility of UK data. [Colour figure can be viewed at wileyonlinelibrary.com]

4.3 | Evidence from simulations

Can the increases in real GDP growth during deflation be aligned with reasonable assumptions on measurement errors? I answer this question using two simulations. First, I construct an error-ridden GDP measure for the 19th century USA, deflating nominal GDP with the CPI and the proxy.[36] This mimics nominal GDP being deflated by an error-ridden CPI based on wholesale prices (see, e.g., Studer & Schuppli, 2008; Kaufmann, 2019; HSSO, 2012, for Switzerland). Second, I simulate data calibrated to match the volatility of GDP growth, inflation, and measurement errors using the modern US replications.[37]

The results are shown in Table 4. If we use the US CPI as a deflator, the results remain qualitatively unchanged. This does not come as a surprise because many of the same data sources are used for constructing the CPI and the GDP deflator, and US real GDP growth relies heavily on industrial production data constructed by Davis (2004). If we use the proxy as

[36] See Online Appendix B, Figure B.3.

[37] I assume that $\alpha = 3$ and $\beta = -1$ and set the standard deviation of inflation to 2.8 (see Table B.4 in Online Appendix B.6). I then construct the deflation indicator based on the simulated inflation series. I choose the variance of the residual ϵ_t in such a way that $y_t = \alpha + \beta x_t + \epsilon_t$ matches the volatility of real GDP growth. Nominal GDP growth is then given by the sum of real GDP growth and inflation. Finally, I simulate three independent measurement error series using the mean and volatility of the last line in Table B.4 in Online Appendix B.6. This provides a simulation with serious measurement problems. I then simulate three error-ridden measures of real GDP. One adds the measurement errors to real GDP directly, introducing i.i.d. measurement errors in the dependent variable. The other two are added to inflation, which is then used to construct the erroneous indicators and deflate well-measured nominal GDP.

TABLE 4 The role of the GDP deflator

	Baseline	Cond. independence			Cond. dependence
US NGDP/CPI:					
$\alpha = E[y\|\pi > 0]$	4.97***	5.54***	5.55***	6.61***	6.54***
	(0.55)	(0.62)	(0.63)	(1.04)	(1.07)
$\beta = E[y\|\pi < 0] - E[y\|\pi > 0]$	−1.25	−2.26**	−2.27***	−4.61**	−3.95**
	(0.87)	(0.89)	(0.87)	(1.94)	(1.55)
US NGDP/proxy:					
$\alpha = E[y\|\pi > 0]$	5.41***	1.76**	1.75**	−3.15	−0.86
	(0.96)	(0.81)	(0.76)	(2.53)	(1.50)
$\beta = E[y\|\pi < 0] - E[y\|\pi > 0]$	−1.05	4.29***	4.30***	16.42***	7.65***
	(1.46)	(1.34)	(1.28)	(5.16)	(2.24)
Simulation:					
$\alpha = E[y\|\pi > 0]$	2.76***	2.88***	2.88***	3.43***	3.09***
	(0.09)	(0.12)	(0.12)	(0.24)	(0.16)
$\beta = E[y\|\pi < 0] - E[y\|\pi > 0]$	−0.53***	−0.83***	−0.83***	−1.76***	−1.19***
	(0.12)	(0.15)	(0.15)	(0.41)	(0.22)
Simulation GDP+i.i.d. error:					
$\alpha = E[y\|\pi > 0]$	2.38***	2.44***	2.44***	3.08***	2.61***
	(0.17)	(0.22)	(0.21)	(0.41)	(0.30)
$\beta = E[y\|\pi < 0] - E[y\|\pi > 0]$	−0.49**	−0.78***	−0.78***	−1.76**	−1.11***
	(0.23)	(0.28)	(0.28)	(0.72)	(0.41)
Simulation NGDP/proxy:					
$\alpha = E[y\|\pi > 0]$	1.22***	1.73***	1.72***	3.46***	1.18***
	(0.15)	(0.20)	(0.20)	(0.45)	(0.28)
$\beta = E[y\|\pi < 0] - E[y\|\pi > 0]$	2.88***	1.87***	1.88***	−1.24	2.71***
	(0.20)	(0.26)	(0.26)	(0.79)	(0.37)
Bound	Upper	Upper	Point	Lower	Point
Method	OLS	OLS	GMM	IV	GMM
Indicator	CPI	CPI, proxy	CPI, proxy	CPI, proxy	CPI, proxy

Note. Model: $y_t = \alpha + \beta x_t + \epsilon_t$ where $x_t \equiv \mathbf{1}_{\{\tilde{\pi}_t<0\}}$. The first two specifications use US GDP growth deflated by either the CPI or the proxy. The estimation sample covers the period 1800–1899. The remaining specifications use 1,000 simulated observations under different assumptions on the measurement errors. Model parameters: Mean growth rate under inflation α; shortfall under deflation β. Baseline: OLS estimates using the CPI. Conditional independence: Bounds and a point estimates using the CPI and proxy, assuming conditional independence. Conditional dependence: Point estimates using the CPI and proxy, assuming that the joint misclassification probabilities equal 0.15 (0.1 for simulated data). HAC-robust standard errors are given in parentheses. ***, **, * denotes significance at the 1%, 5%, 10% level.

a deflator we find a significantly positive association between deflation and real GDP growth, at least when exploiting multiple indicators.[38]

The simulation exercise qualitatively confirms these results. The true coefficients are set to $\alpha = 3$ and $\beta = -1$. We see that i.i.d. measurement errors in GDP growth only increase the standard errors. However, if we use the error-ridden indicator to deflate well-measured nominal GDP, the coefficients switch sign and are of the same magnitude as those we find for Switzerland, Finland, and Spain.

4.4 | Robustness

I conducted a range of robustness tests using US data, which are shown in Online Appendix D.

The literature suggests that asset price declines matter more than CPI deflation (see, e.g., Borio et al., 2015). In addition, Jalil (2015) shows that the US price level declined significantly after major banking panics. Finally, deflation and banking crises may be caused by tight monetary policy (see, e.g., Friedman & Schwartz, 1963, p. 108). I therefore add dummies for stock price declines, banking crises, and slow growth in a monetary aggregate.[39] Controlling for additional covariates

[38] That multiple indicators exacerbate the bias if the measurement errors are correlated is in line with additional cross-country evidence reported in Online Appendix C. For some of the countries I obtained a second price measure. The relationship between the OLS coefficients and the volatility of real GDP growth and inflation becomes more pronounced.

[39] When adding covariates the model is over-identified so that we can test the assumptions underlying the model (see Online Appendix A). We do not reject the over-identifying restrictions in any of the specifications.

leaves the association between deflation and real activity intact. The bounds, as well as the GMM point estimates, suggest that deflation is associated with significantly lower industrial production growth.

The results are robust to accounting for additional deflation indicators. I obtained an alternative price index developed by Falkner for the period 1840–1891 (Aldrich Report, 1893 p. 93).[40] This price index is not included in the composite CPI constructed by Officer and Williamson (2016) because of its many deficiencies.[41] The individual estimates suggest a shortfall in industrial production growth of −2.1 pp to −5.1 pp. The estimate is not statistically significant using only the Falkner index. Combining two indicators always yields a significantly negative association. If we use all three indicators, we find an upper bound for the shortfall in industrial production growth of −5.5 pp. This suggests that using the Falkner index instead of the proxy yields qualitatively similar results.

I also varied the definition of the deflation classification because some deflationary periods may be more harmful than others (see Bordo & Filardo, 2005). When estimating the association for severe deflations, that is, price declines of more than 3%, the OLS estimate becomes more negative and statistically significant. The conservative upper bound still suggests that industrial production growth declines by at least 1 pp more. The results are also robust when focusing on persistent deflations of two or more years. The OLS estimate amounts to −1.6 pp and is not statistically significant. Accounting for measurement errors, we find a significant shortfall in industrial production growth of between −3.9 pp and −6.9 pp.

To test whether the results depend on the identifying assumptions I estimate the model for all combinations of the misclassification rates between 0% and 15% in steps of 5%. Most estimates point to a similar shortfall during periods of deflation. In addition, the biases are mostly statistically significant.

The estimators assume that the error term is i.i.d. If this assumption is violated the GMM estimate is biased (see Cosslett & Lee, 1985). Black et al. (2000) suggest that the bounding strategy still works if the measurement errors are uncorrelated with other covariates. I therefore estimate the model by Black et al. (2000) with one lag of the dependent variable as an additional covariate.[42] The results do not change because US industrial production growth shows little persistence during the 19th century.

5 | CONCLUSION

Many empirical studies using 19th century data find no significant link between real economic activity and deflation. This is possibly because deflationary episodes were benign, short-lived, or a consequence of beneficial advances in productivity. This paper proposes an additional explanation: measurement errors in historical price data.

Is deflation costly after all? This paper suggests that deflation is at least more costly than previously thought. Although the results cannot be directly extrapolated to the present, the findings have contemporary implications: The 19th century evidence does not lead to the conclusion that policy makers' fear of deflation should be dismissed.

In addition, the findings have implications for a rapidly growing literature using historical cross-country data (see, e.g., Reinhart & Rogoff, 2010; Schularick & Taylor, 2012; Jordà et al., 2016; Baker et al., 2018). Importantly, adding an additional country may bias the results if the data quality is poor. In addition, if measurement errors are correlated across multiple series, the direction and severity of the biases are difficult to anticipate without detailed knowledge of the underlying data. Ultimately, the biases may well offset the efficiency gains from adding more countries.

Whether measurement errors in historical price data distort structural analyses remains an open question. Accurately estimating reduced-form correlations, however, is a first step towards improving our understanding of the 19th century economy. Examining the impact of measurement errors in price data on structural analysis, for example along the lines of Bayoumi and Eichengreen (1996), Bordo and Redish (2004) and Beckworth (2007), is an interesting avenue for future research.

ACKNOWLEDGEMENT

This project benefited from a visit to the Berkeley Economic History Laboratory (BEHL), whose hospitality I gratefully acknowledge. I thank the editor, Marco Del Negro, Christiane Baumeister, Gregor Bäurle, Bernd Bartels, Gillian Brunet, Brad DeLong, Barry Eichengreen, Stefan Gerlach, Yuriy Gorodnichenko, Savina Gygli, Patrick Halbeisen, Philipp Harms,

[40] Falkner applies consumer expenditure weights to wholesale prices. However, he uses different data sources than included in the proxy. There is small overlap for five fruit items with the index constructed by Hoover (1960).

[41] See Hoover (1960) and Long (1960) for a description of the most important deficiencies.

[42] See Online Appendix A for a discussion.

Bo Honoré, Matthias Hölzlein, Florian Huber, Ronald Indergand, Jan Jacobs, Dmitri Koustas, Carlos Lenz, Ronan Lyons, Petros Milionis, Andreas Müller, Klaus Neusser, Nuno Palma, Steve Pischke, Tobias Renkin, Christina Romer, Gisela Rua, Hugo Saner, Samad Sarferaz, Jan-Egbert Sturm, Michael Siegenthaler, Eric Sims, Zach Stangebye, Rebecca Stuart, 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, the WU Vienna, the KOF/ETH Zurich, the SSES Annual Meeting, and the Economic History Workshop of the Central Bank of Ireland 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.

CONFLICT OF INTEREST STATEMENT

I hereby declare no conflict of interest.

OPEN RESEARCH BADGES

This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at [http://qed.econ.queensu.ca/jae/datasets/kaufmann004/].

REFERENCES

Adams, T. M. (1939). Prices paid by farmers for goods and services and received by them for farm products, 1790-1871; Wages of farm labor, 1789-1937: A preliminary report. Burlington, VT: University of Vermont and State Agricultural College.

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J., & Li, H. (2019). Missing growth from creative destruction. *American Economic Review*, *109*(8), 2795–2822.

Aigner, D. J. (1973). Regression with a binary independent variable subject to errors of observation. *Journal of Econometrics*, *1*, 49–60.

Aldrich Report (1893). Wholesale prices, wages, and transportation. Retrieved from hdl.handle.net/2027/chi.20247016

Allen, S. G. (1992). Changes in the cyclical sensitivity of wages in the United States, 1891-1987. *American Economic Review*, *82*(1), 122–140.

Atkeson, A., & Kehoe, P. J. (2004). Deflation and depression: Is there an empirical link? *American Economic Review*, *94*(2), 99–103.

Baker, S. S., López-Salido, J. D., & Nelson, E. (2018). The money view versus the credit view. Finance and Economics Discussion Series *2018-042*): Board of Governors of the Federal Reserve System.

Barsky, R. B. (1987). The Fisher hypothesis and the forecastability and persistence of inflation. *Journal of Monetary Economics*, *19*(1), 3–24.

Bayoumi, T., & Eichengreen, B. (1996). The stability of the Gold Standard and the evolution of the International Monetary Fund. In Bayoumi, T., Eichengreen, B., & Taylor, M. P. (Eds.), *Modern Perspectives on the Gold Standard*. (pp. 165–188). Cambridge: Cambridge University Press.

Bernanke, S. B. (2010). The Economic Outlook and Monetary Policy: Federal Reserve Board www.federalreserve.gov/newsevents/speech/bernanke20100827a.htm

Beckworth, D. (2007). The postbellum deflation and its lessons for today. *The North American Journal of Economics and Finance*, *18*(2), 195–214.

Benati, L. (2008). Investigating inflation persistence across monetary regimes. *The Quarterly Journal of Economics*, *123*(3), 1005–1060.

Black, D. A., Berger, M. C., & Scott, F. A. (2000). Bounding parameter estimates with nonclassical measurement error. *Journal of the American Statistical Association*, *95*(451), 739–748.

Bordo, M. D., & Filardo, A. (2005). Deflation and monetary policy in a historical perspective: Remembering the past or being condemned to repeat it? *Economic Policy*, *20*(44), 799–844.

Bordo, M. D., & Redish, A. (2004). Is deflation depressing? Evidence from the classical Gold Standard. In Burdekin, R. C. K, & Siklos, P. L. (Eds.), Deflation: Current and Historical Perspectives (pp. 99–202). Cambridge: Cambridge University Press.

Borio, C., Erdem, M., Filardo, A., & Hofmann, B. (2015). The costs of deflations: A historical perspective. *BIS Quarterly Review*, *2015*, 31–54.

Boskin Commission (1996). Toward a more accurate measure of the cost of living: Final report to the Senate Finance Committee from the Advisory Committee to study the consumer price index. Retrieved from www.ssa.gov/history/reports/boskinrpt.html

Bound, J., & Krueger, A. B. (1991). The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *Journal of Labor Economics*, *9*(1), 1–24.

Card, D. (1996). The effect of unions on the structure of wages: A longitudinal analysis. *Econometrica*, *64*(4), 957–979.

Cogley, T., & Sargent, T. J. (2015). Measuring price-level uncertainty and instability in the United States, 1850-2012. *The Review of Economics and Statistics*, *97*(4), 827–838.

Cosslett, S. R., & Lee, L.-F. (1985). Serial correlation in latent discrete variable models. *Journal of Econometrics*, *27*(1), 79–97.

David, P. A., & Solar, P. (1977). Research in economic history: Vol. 2. *A bicentenary contribution to the history of the cost of living in America, Research in Economic History*, Vol. *2*. Greenwich: JAI Press.

Davis, J. H. (2004). An annual index of US industrial production, 1790-1915. *The Quarterly Journal of Economics*, *119*(4), 1177–1215.

Eichengreen, B., Park, D., & Shin, K. (2016). Deflation in Asia: Should the dangers be dismissed? (ADB Economics Working Paper Series 490) Mandaluyong, Philippines: Asian Development Bank.

Friedman, M., & Schwartz, A. (1963). *A monetary history of the united states, 1867–1960*. Princeton, NJ: Princeton University Press.

Goolsbee, A. D., & Klenow, P. J. (2018). Internet rising, prices falling: Measuring inflation in a world of e-commerce. *AEA Papers and Proceedings*, *108*, 488–92.

Griliches, Z. (1986). Economic data issues, In Griliches, Z., & Intriligator, M. D. (Eds.) *Handbook of econometrics* (Vol. 3, pp. 1465–1514). Amsterdam: Elsevier.

HSSO (2012). Nationale Buchhaltung: Tab. Q.1a., Historische Statistik der Schweiz. Retrieved from www.hsso.ch/2012/q/

Hanes, C. (1998). Consistent wholesale price series for the United States, 1860-1990. In Dick, T. J. O. (Ed.), (*Business cycles since 1820: New international perspectives from historical evidence*. pp.176–209). Cheltenham: Edward Elgar.

Hartmann, P., & Smets, F. (2018). The first twenty years of the European Central Bank: Monetary policy (Working Paper Series 2219). Frankfurt, Germany: European Central Bank.

Hausman, J. (2001). Mismeasured variables in econometric analysis: Problems from the right and problems from the left. *The Journal of Economic Perspectives*, *15*(4), 57–67.

Hoover, E. D. (1958). Wholesale and retail prices in the nineteenth century. *The Journal of Economic History*, *18*(03), 298–316.

Hoover, E. D. (1960). Retail prices after 1850, Trends in the American economy in the nineteenth century (pp. 141–190). Princeton, NJ: Princeton University Press.

Jalil, A. J. (2015). A new history of banking panics in the United States, 1825-1929: Construction and implications. *American Economic Journal: Macroeconomics*, *7*(3), 295–330.

Johnston, L., & Williamson, S. H. (2019). Sources and techniques used in the construction of annual GDP, 1790-1928. Note, MeasuringWorth, retrieved from www.measuringworth.com/usgdp/

Jordà, Ò., Schularick, M., & Taylor, A. M. (2016). Macrofinancial history and the new business cycle facts, *NBER Macroeconomics Annual 2016*, vol. 31 Chicago, IL: University of Chicago Press.

Kane, T. J., Rouse, C. E., & Staiger, D. (1999). Estimating returns to schooling when schooling is misreported (NBER Working Papers 7235) Cambridge, MA: National Bureau of Economic Research.

Kaufmann, D. (2019). Nominal stability over two centuries. *Swiss Journal of Economics and Statistics*, *155*, 1–23.

Kreider, B. (2010). Regression coefficient identification decay in the presence of infrequent classification errors. *The Review of Economics and Statistics*, *92*(4), 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*. Princeton: Princeton University Press.

Mahajan, A. (2006). Identification and estimation of regression models with misclassification. *Econometrica*, *74*(3), 631–665.

Margo, R. A. (2000). *Wages and labor markets in the united states, 1820-1860*. Chicago: University of Chicago Press.

Officer, L. H. (2014). What was the consumer price index then? A data study. Note, MeasuringWorth, retrieved from www.measuringworth.com/docs/cpistudyrev.pdf

Officer, L. H., & Williamson, S. H. (2016). Annual consumer price index for the United States, 1774-2015 Note, MeasuringWorth, retrieved from www.measuringworth.com/uscpi/

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*(2), 491–497.

Reinhart, C. M., & Rogoff, K. S. (2010). Growth in a time of debt. *American Economic Review*, *100*(2), 573–578.

Ritzmann, H., & David, T. (2012). Schätzung des Bruttoinlandprodukts nach Branchen und Kantonen 1890-1960.

Romer, C. (1986a). Is the stabilization of the postwar economy a figment of the data? *American Economic Review*, *76*, 314–3347.

Romer, C. (1986b). Spurious volatility in historical unemployment data. *Journal of Political Economy*, *94*(1), 1–37.

SNB (2011). Swiss National Bank sets minimum exchange rate at CHF 1.20 per Euro: Swiss National Bank www.snb.ch/en/mmr/reference/pre_20110906/source/pre_20110906.en.pdf

Shoemaker, O. J. Variance estimates for price changes in the consumer price index: January-December 2014. US Bureau of Labour Statistics, retrieved from www.bls.gov/cpi/tables/variance-estimates/home.htm

Schularick, M., & Taylor, A. M. (2012). Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870-2008. *American Economic Review*, *102*(2), 1029–1061.

Studer, R., & Schuppli, P. (2008). Deflating Swiss prices over the past five centuries. *Historical Methods*, *41*(3), 137–153.

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

Weeks, J. D. (1886). Report on the statistics of wages in manufacturing industriesWashington, DC: Department of the Interior, Census Office. hdl.handle.net/2027/hvd.hl4p9r.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Kaufmann D. Is deflation costly after all? The perils of erroneous historical classifications. *J Appl Econ*. 2020;35:614–628. https://doi.org/10.1002/jae.2762