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

Daniel Kaufmann[†,‡]

13 November 2019

Abstract: This paper estimates economic activity during deflationary and inflationary periods accounting for measurement errors in 19th century price data. Because of measurement errors, we underestimate (overestimate) economic activity during inflation (deflation). With multiple erroneous deflation indicators we can recover the true relationship. After accounting for measurement errors, the shortfall of U.S. industrial production growth during deflation ranges from -4.5 pp to -7.6 pp, instead of -2 pp. In addition, there is a negative association between deflation and various real activity measures in the U.K. Finally, I provide estimates of average GDP per capita growth during deflation for eleven countries. The cross-country variation in the estimates is in line with the idea that the estimates are more distorted for countries with more serious measurement errors in price data.

JEL classification: E31, E32, N11, C25

Keywords: Deflation, real activity, 19th century, measurement error, binary regressors, misclassification bias, deflator bias, bounds, GMM.

1 Introduction

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

So 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 a dependent variable and a binary deflation indicator as an independent variable.[1] If this binary indicator is measured with error, OLS fails to recover the true coefficients, for two reasons. First, OLS suffers from a misclassification bias (see Aigner 1973). If we use a mismeasured price index to classify inflationary (deflationary) episodes, some of these episodes were actually associated with falling (rising) prices and therefore low (high) real activity. Second, OLS suffers from a deflator bias if the GDP deflator has similar deficiencies as the price index underlying the classification. Some of the inflation (deflation) episodes stem from positive (negative) measurement errors in the price index, which lead to negative (positive) measurement errors in real GDP growth.

I propose a strategy to resolve these biases. First, I use multiple error-ridden binary deflation indicators for the 19th century U.S. to resolve the misclassification bias. The estimator builds on Kane et al. (1999) and Black et al. (2000) who deal with misreported binary responses in survey data. Their estimator requires measurement issues in the binary indicators are unrelated.[2] However, most historical price indices suffer from similar measurement problems. I therefore propose an estimator allowing for correlated measurement problems assuming 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 deal with the deflator bias, I focus the

---

[1] The regression framework allows to quantitatively assess and resolve the bias. Similar issues arise in qualitative analyses conditional on a binary classification in deflation and inflation periods (see e.g. Friedman and 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 second indicator has to be uncorrelated with the measurement errors in the first.

main analysis on real activity data that does not depend on a deflator (industrial production and unemployment) and on countries with arguably well measured 19th century data (U.S. and U.K.).

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 U.S. 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, I find a significant shortfall in U.S. GDP growth, as well as a negative impact on U.K. industrial production, unemployment, consumption, and investment.

Why do these results differ from the existing literature? To shed some light on this question, I analyzed GDP per capita growth for eleven countries. The varying quality of the price data among these countries introduces varying degrees of the deflator bias. Indeed, 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 deflation. These patterns provide suggestive evidence for the deflator bias. If large measurement errors in prices introduce negatively correlated errors in real GDP, we overestimate growth during deflation more for those countries with more serious measurement problems.

This paper is related to studies on the real effects of deflation. Because prices frequently declined under metallic standards, most studies use 19th century data which suffer from more serious measurement problems 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 on average over a large number of countries a weak link between real activity and deflation. Eichengreen et al. (2016) report the link becomes stronger when using wholesale prices instead of consumer prices. One explanation for the weak link is that many 19th century deflations were driven by advances in productivity (see Friedman and 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 and Eichengreen 1996, and references therein). This paper argues in favour of a third possibility: measurement errors in 19th century price data biases estimates of economic growth during deflation.

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 and Sargent 2015; Eichengreen et al. 2016).[3] Others examine the robustness of the results using alternative price indices (see Margo 2000, ch. 2). I argue that all 19th century price indices are imperfect measures. Therefore, we can jointly exploit these indices to improve estimates of economic growth during deflation.

The paper is also related to Romer (1986a,b), Allen (1992), and Hanes (1998). They apply 19th century methodologies to modern data to assess the properties and consequences 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 a historical analysis. 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 fix the misclassification rate in reported union status in an analysis of wages from 1987–1988. Also, Bound and Krueger (1991) use an earnings survey matched with administrative data to examine the severity and properties of reporting error.

There is an extensive literature showing measurement problems account for missing growth in modern data. Modern price data overestimate inflation for several reasons and, as a result, we underestimate economic growth. The Boskin Commission (1996) showed that missing quality adjustments in modern CPI data causes 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. (2017) suggest that disappearing products introduce an upward bias in inflation and, therefore, a downward bias in economic growth. Similarly, this paper argues measurement problems account for the missing shortfall of economic activity during deflation.

The remainder of the paper is structured as follows. I first discuss the properties of the misclassification and deflator biases and potential remedies. Then, I discuss the properties of the misclassification error in 19th century price data using modern replications. These data then allow to consistently estimate economic activity during inflation and deflation periods in the 19th century.

---

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

The last section concludes.

2 Misclassification and deflator biases

Measurement errors in historical data are often regarded to attenuate the correlation between macroeconomic times series.[4] The intuition follows from the 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). This intuition fails in two ways when assessing the relationship between real activity and deflation. First, because the mismeasured independent variable is a binary dummy, the measurement errors are necessarily correlated with the true variable. Second, if the GDP deflator suffers from similar deficiencies as the price index used to classify deflation periods, the measurement errors of the dependent variable will be correlated with the measurement errors in the independent variable.

This section describes the properties of the misclassification and deflator biases both analytically and in a simulation exercise. I then propose various estimation approaches to recover the true relationship between real activity and deflation.

2.1 Measurement errors in real activity and inflation

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:[5]

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

where y_t represents real activity, $d_t \equiv \mathbf{1}_{\{\pi_t < 0\}}$ a binary deflation indicator, and ε_t an i.i.d. error term. A negative β implies deflation is associated with a decline in real activity. A positive α implies 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 cannot consistently estimate Eq. (1) by OLS.

---

[4]See, for example, Jordá et al. (2016): "We also confirm that consumption and investment are procyclical with output. This comovement seems to increase over time, potentially reflecting better measurement."

[5]See Borio et al. (2015), Eichengreen et al. (2016), Reinhart and Rogoff (2010), and Jordá et al. (2016) for historical studies showing descriptive statistics on various economic relationships that can be cast in this model.

Instead, we estimate:

$$\tilde{y}_t = \alpha + \beta x_t + \epsilon_t \tag{2}$$

$$\epsilon_t = \varepsilon_t - \beta(x_t - d_t) + (\tilde{y}_t - y_t),$$

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 ($x_t - d_t$). If these errors were i.i.d., the OLS estimate of β would suffer from the attenuation bias. However, the classification error is necessarily correlated with the true binary dummy d_t (see Aigner 1973).[6] In addition, the error term is correlated with the erroneous binary dummy if the price index used to deflate real activity suffers from similar deficiencies as the index used to classify deflation.

2.2 Properties of the biases

To characterize the direction and size of the biases, I derive the probability limit of the OLS estimator analytically, and examine its properties in a simulation exercise.

The probability limit of the OLS estimator equals the difference in expectations, conditional on each outcome of the binary dummy (see Aigner 1973). Assuming the measurement errors in real activity are unrelated to the measurement errors in inflation:

$$plim\ \hat{\alpha}_{ols} = E[\tilde{y}_t | x_t = 0] \tag{3}$$

$$= \alpha + \beta P[d_t = 1 | x_t = 0]$$

$$plim\ \hat{\beta}_{ols} = E[\tilde{y}_t | x_t = 1] - E[\tilde{y}_t | x_t = 0]$$

$$= \beta(1 - P[d_t = 0 | x_t = 1] - P[d_t = 1 | x_t = 0]).$$

If $\beta < 0$ we will underestimate real activity during inflationary periods ($plim\ \hat{\alpha}_{ols} < \alpha$). In addition, we will overestimate the short-fall in real activity during deflation ($plim\ \hat{\beta}_{ols} > \beta$). In other words,

---

[6]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.

deflationary periods appear too benign, and inflationary periods less buoyant.

The bias becomes larger if the measurement errors in real activity are negatively correlated with the measurement errors in inflation. Suppose we measure nominal GDP growth without error (n_t).[7] But, 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 of real GDP growth are negatively correlated with the measurement errors in inflation $\tilde{y}_t - y_t = n_t - \tilde{\pi}_t - (n_t - \pi_t) = -(\tilde{\pi}_t - \pi_t)$. The probability limit of the OLS estimator then equals:

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

The bias becomes larger 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 at the same time lead to lower (higher) real GDP growth because of the error-ridden deflator.

To assess the properties and severity of the biases, I simulate the probability limit of $\hat{\alpha}_{ols}$ and $\hat{\beta}_{ols}$ assuming that the error-ridden inflation measure depends linearly on actual inflation, distorted by an i.i.d. error (ω_t):

$$
\tilde{\pi}_t = \rho_0 + \rho_1 \pi_t + \omega_t\ , \qquad\qquad (5)
$$

The functional form allows for a mismeasured intercept (ρ_0) and slope (ρ_1). In addition, the binary indicator depends on a threshold c ($x_t \equiv \mathbf{1}_{\{\tilde{\pi}_t < c\}}$).[8] In the baseline, I examine the impact of i.i.d. measurement errors ($\rho_0 = 0$, $\rho_1 = 1$, $c = 0$). 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.[9] Then, I vary the threshold ($c = 5$), the intercept ($\rho_0 = 5$), and the slope parameter ($\rho_1 = 3$). In addition, I examine the

---

[7]Allowing for i.i.d. measurement errors in nominal GDP would not alter the derivation of the probability limits.

[8]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%.

[9]The overall volatility corresponds to inflation volatility in the 19th century U.S.

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 measurement errors in real activity are unrelated to measurement errors in inflation. Panels (c) and (d) show the probability limit with correlated measurement errors. The horizontal dotted lines give the true values of the parameters.[10]

Five observations emerge. First, the misclassification bias pushes the the probability limits of the coefficients in opposite directions (dashed lines in panels a and b).[11] But the bias in β is larger in absolute size, so we will overestimate real activity during deflation. Second, the bias becomes larger when increasing the threshold c (solid line). Choosing a threshold that differs from the true unconditional average implies we classify a rare event. It therefore becomes more likely that the indicator misclassifies the event.[12] Third, a mismeasured intercept exacerbates the bias (dashed-dotted line). 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 (dotted line). A larger slope increases the probability that the indicator classifies an episode correctly. 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, unless the signal-to-noise ratio becomes very large.

These results have implications for studies analyzing the link between deflation and real activity. For countries with more serious measurement problems, we are likely to underestimate (overestimate) real activity during inflation (deflation) episodes. Moreover, measurement errors in the mean biases the OLS estimate.[13] This bias is likely relevant because, even today, we overestimate inflation because of technological progress, creative destruction, and product substitution (Boskin Commission 1996;

---

[10]Online Appendix C, Figure C.1., shows simulations for the growth during deflation ($\alpha + \beta$).

[11]Note that 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 misclassfification can lead to even more serious biases.

[12]The same phenomenon occurs in diagnosing rare illnesses. A small rate of false positives may imply that most of the positively tested individuals are in fact healthy.

[13]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

(a) α with mismeasured inflation

(b) β with mismeasured inflation

(c) α with mismeasured real activity and inflation

(d) β with mismeasured real activity and inflation
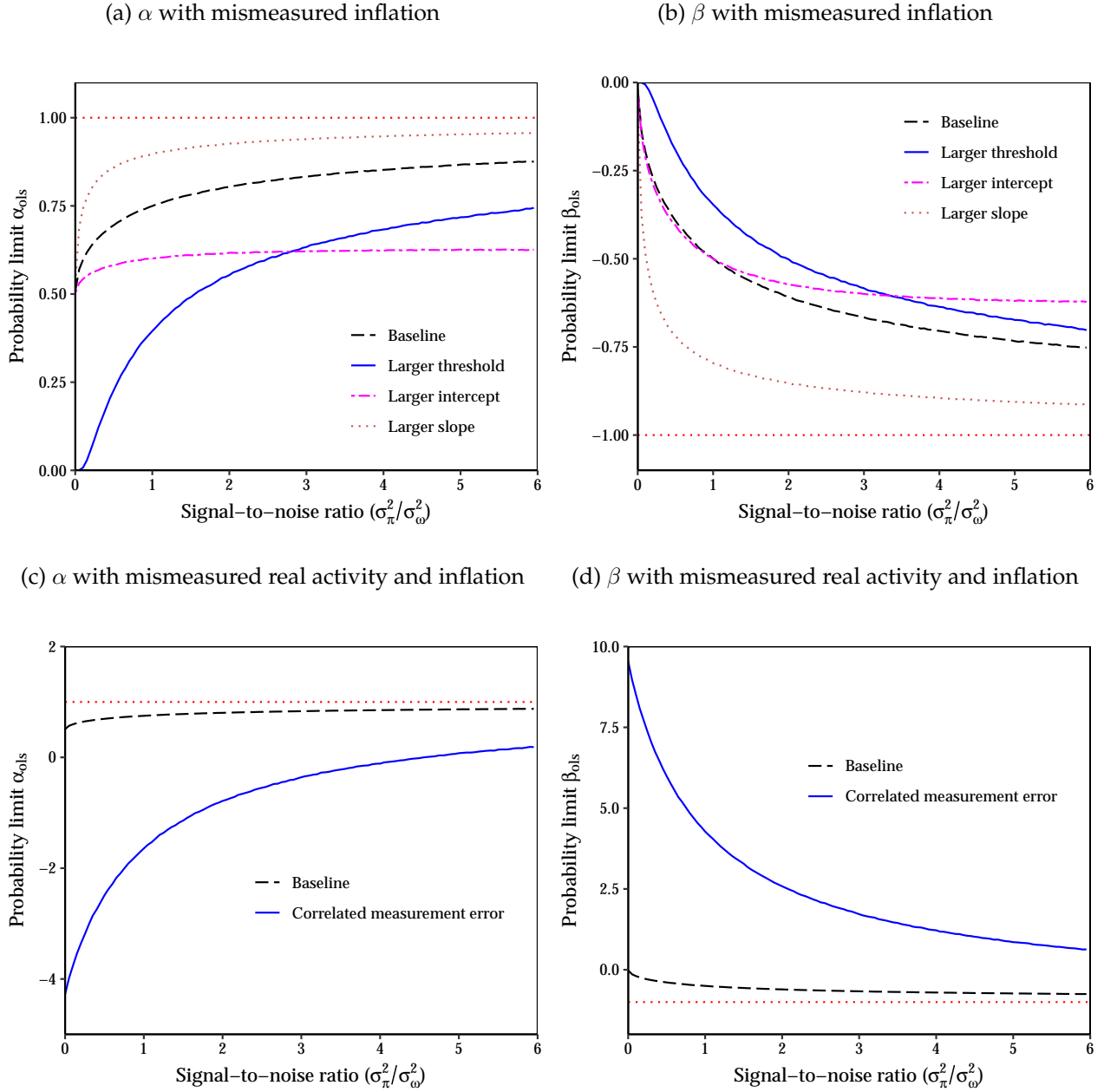
Notes: The figure shows the probability limit of the OLS estimator of $y_t = \alpha + \beta x_t + \epsilon_t$, where $x_t \equiv \mathbf{1}_{\{\tilde{\pi}_t < c\}}$, as a function of the signal-to-noise ratio ($\sigma_\pi^2/\sigma_\omega^2$). The dotted horizontal lines give the true value of $\alpha = 1$ 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 $c = 0$, $\rho_0 = 0$, $\rho_1 = 1$, and $\sigma = \sqrt{\sigma_\pi^2 + \sigma_\omega^2} = 6$ (dashed line). 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 Eq. 3). Panels (c) and (d) show the probability limit with negatively correlated measurement errors in the dependent variable (see Eq. 4).

Aghion et al. 2017; Goolsbee and Klenow 2018). Then, the bias is more pronounced if we examine rare events, for example, particularly severe deflations. The bias is mitigated, however, if we obtain an inflation measure that puts more weight on volatile, but well-measured, prices. This could be one reason for the different results found by Borio et al. (2015) using consumer prices and Eichengreen et al. (2016) using wholesale prices. Finally, the deflator bias strongly distorts the probability limit. This suggests that we should use real activity measures that are directly estimated in real terms, rather than using real GDP, which may suffer from correlated measurement errors through the deflator.

2.3 Bounds and consistent estimators

We can resolve the misclassification bias if we obtain a second binary indicator ($z_t \equiv \mathbf{1}_{\{\hat{\pi}_t < 0\}}$) based on a second error-ridden inflation rate ($\hat{\pi}_t$). This section outlines the estimators.[14] I propose various strategies assuming that measurement errors in real activity and inflation are unrelated. To address the deflator bias, the empirical analysis uses various real activity measures, which are directly estimated in real terms.

2.3.1 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:

$$ y_t = \alpha + \beta_{11}\mathbf{1}_{\{x_t=1, z_t=1\}} + \beta_{10}\mathbf{1}_{\{x_t=1, z_t=0\}} + \beta_{01}\mathbf{1}_{\{x_t=0, z_t=1\}} + \epsilon_t \; , \tag{6} $$

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

In addition, we can use the second proxy as an instrumental variable (see Hausman 2001). Kane et al. (1999) emphasize that IV does not resolve the bias because the misclassification error is necessarily

---

[14]See Online Appendix A for more details.

correlated with the true regressor. The probability limit of the IV estimator equals:

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

IV yields a lower bound if $\beta < 0$ because the misclassification probabilities are zero or positive.

Finally, we can consistently estimate α and β using GMM (see Kane et al. 1999; Black et al. 2000, and Online Appendix A). Note that we can estimate seven empirical moments from the data: three sampling fractions of the binary indicators, as well as, 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 if we would only use one indicator, as a non-linear function of the underlying GMM estimates.

2.3.2 Conditional dependence

In the present application the conditional independence assumption is probably too strong.[15] 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.[16] When the measurement errors in the inflation data are only weakly correlated, these probabilities are small. To provide additional evidence on the size of these probabilities, the next section explores the properties of measurement errors in historical price data for the U.S. using modern replications.

---

[15]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.

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

3 Measurement errors in 19th century data

To consistently estimate the shortfall in real activity during deflation we need multiple deflation indicators, as well as, information on the joint misclassification probabilities. In addition, we need real activity data that is unrelated to the price indices used in the classification. This section presents such data for the U.S. and discusses the most important sources and properties of measurement errors.

3.1 CPI data

The first price index used in the empirical analysis is the composite CPI of Officer and Williamson (2016). The series likely represents the most accurate CPI at a given point in time. Nevertheless, the pre-20th century segments suffer from various methodological deficiencies.

Table 1 — Methodological deficiencies

Deficiency	Source	Time span	Comments
Wholesale prices	David and Solar (1977)	1774-1851	Wholesale prices approximate retail prices before 1800
Geographical coverage	David and Solar (1977)	1774-1851	Prices for Philadelphia (before 1800) and prices paid by Vermont farmers (until 1851)
Sample size	Hoover (1960)	1851-1860	Weeks Report shows many missing observations and small number of individual price quotes
	Long (1960)	1880-1890	Little information on retail prices for the entire decade after the Weeks Report ends
Reproduction cost index	Lebergott (1964)	1860-1880	Rent approximated by prices of construction materials and wages of low-skilled workers
Few services	Lebergott (1964)	1860-1880	
Linear interpolation	Long (1960)	1880-1890	Several items interpolated over the entire decade (particularly rent)

Notes: The time span represents the segment used in the composite CPI by Officer and Williamson (2016) reported in Officer (2014).

Table 1 gives an overview. Before 1800, David and Solar (1977) use wholesale prices to approximate prices at the retail stage. 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 stem from farmers in Vermont. Afterwards, 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. Hoover (1958) emphasizes retail price data are particularly scarce from 1880 to 1890, after the Weeks Report ends and before the U.S. Bureau of Labor Statistics starts 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. 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 by Lebergott (1964) and Hoover (1960) comprise only few service items: rent, shoe repairs, and physician fees.[17]

3.2   A new proxy

The second price index is a newly constructed 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 (see Online Appendix B).[18] Using wholesale prices has three advantages: The data stem from other sources and therefore the measurement errors of the two indicators are to some extent independent;[19] the underlying data is accurate because the information stems from newspapers and company archives (see Hoover 1958); misclassification may be less severe because wholesale prices fluctuate more strongly than consumer prices.

The upper chart in Figure 2 shows 19th century inflation measured by the CPI and the proxy. The mean of the two series is identical (-0.2%). In addition, they display similar turning points and the correlation is sizeable (0.74). However, because the proxy uses wholesale prices, the standard deviation is higher compared to the CPI (9.9 compared to 6.2). Despite their high correlation, the two price series give a different classification of inflation or deflation in 29% of the observations.

---

[17]Similar issues plague CPI estimates well into the 20th century. The U.S. CPI was published on a monthly basis since 1940; even then, many service prices were still collected quarterly. Before, the CPI was only published for irregular intervals or even only for December. From 1913-1921, the BLS retrospectively estimated a monthly CPI, interpolating prices for many items that were not collected monthly (see Officer 2014).

[18]The proxy covers more than 70% of a 19th century consumption basket. The most important missing item is rent (18% of the consumption basket).

[19]Ideally, the errors of the CPI and the proxy are independent. A necessary condition is that the underlying data sources differ, which is the case for most of the 19th century. In Online Appendix B, I show that this is indeed the case for most of the 19th century.

Figure 2 — CPI inflation, proxy, and modern replications
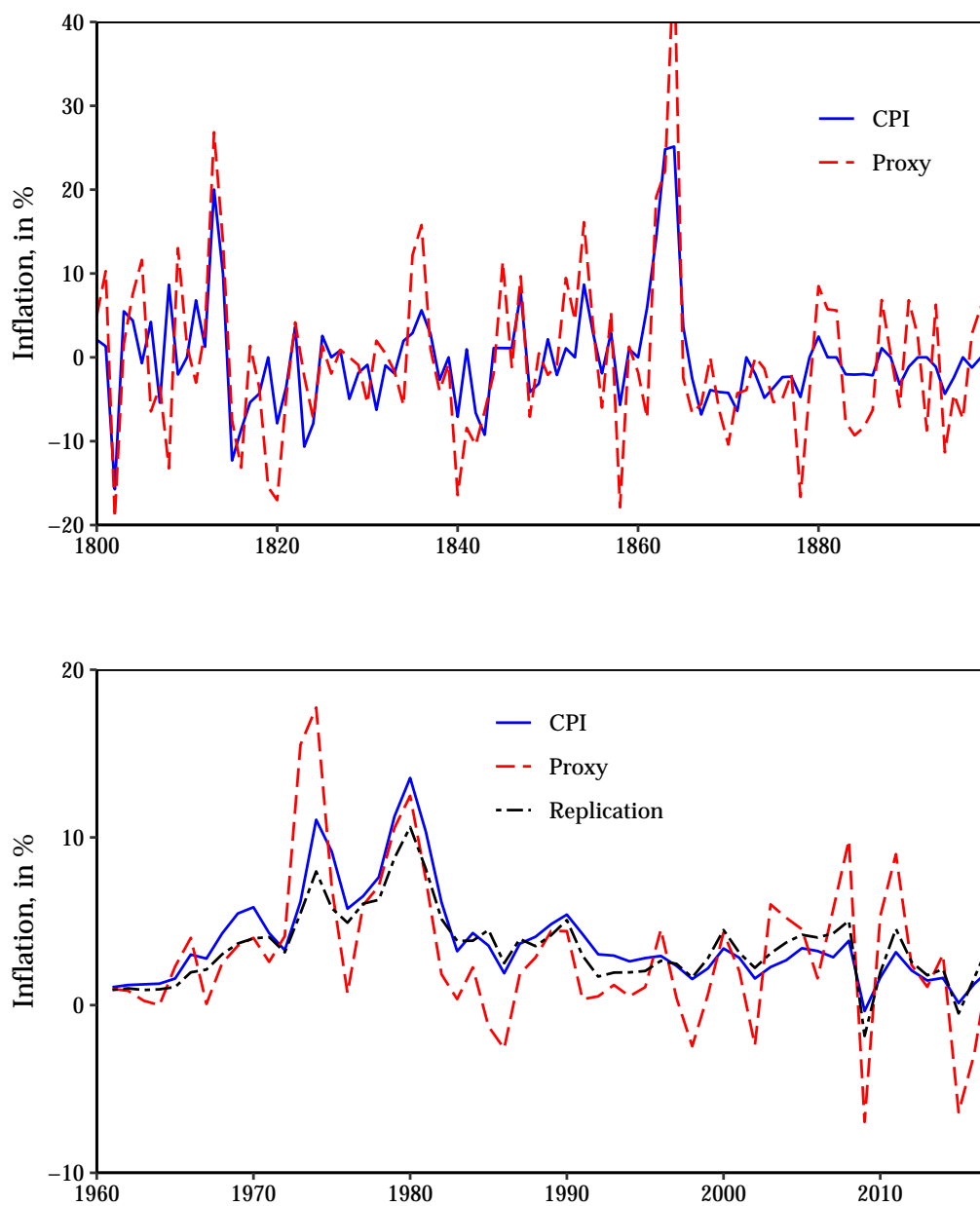
Notes: The proxy combines wholesale and producer price indices with consumer expenditure weights from Hoover (1960) for the 19th century, as well as CPI weights for 2017 from the BLS for the post-WWII period. CPI inflation stems from Officer and Williamson (2016). The replication uses modern data to reconstruct the 19th century CPI by Hoover (1960).

3.3 Evidence from modern replications

Despite distinct data sources, the measurement errors affecting the CPI and the proxy are likely to be correlated because both indices suffer from similar deficiencies. For example, information on services is scarce in both price indices. If this is the case, 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 as well as the well-measured series, we can gauge the joint misclassification rates and the correlation between measurement errors.[20]

I replicate the 19th century CPI by Hoover (1960) using modern data for the period 1960-2017 (Online Appendix B provides a detailed description). In addition, I construct the wholesale price proxy for the post-WWII era using modern PPI data from the U.S. Bureau of Labor Statistics. The lower chart 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 proxy is more volatile than the actual CPI inflation and the replication. 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 for a discussion).

Table 2 — Descriptive statistics (1960-2017)

Statistic	Mean	St. Dev.	Min	Max
CPI	3.82	2.82	−0.36	13.55
Proxy	3.07	4.50	−6.97	17.75
Replication	3.47	2.20	−1.84	10.62
CPI-replication	0.35	1.14	−1.63	3.37
CPI-proxy	0.75	3.11	−9.35	6.61

Notes: Inflation rates measured in percent, differences measured in percentage points. All statistics calculated using annual data from 1960-2017.

Table 2 shows descriptive statistics of the various inflation measures, as well as for the corresponding

---

[20]That the modern CPI is measured without error is not exactly right. The Boskin Commission (1996) shows that the CPI may underestimate actual CPI inflation because of neglected changes in quality and product substitution (see Goolsbee and Klenow 2018, for a more recent study). However, the sampling error in modern CPIs is small (see Shoemaker 2014).

errors, from 1960–2017. The errors are on average positive, suggesting that the proxy and replication on average underestimate inflation. The standard deviation of CPI inflation amounts to 2.8. Meanwhile, the standard deviation of the errors ranges from 1.1–3.1. Therefore, the signal-to-noise ratio ranges from 0.8–6.1. Even for a relatively high signal to noise ratio, Figure 1 suggests we may find a relevant bias.

Table 3 — Properties of measurement errors (1960-2017)

	Replication	Proxy	CPI-replication	CPI-proxy
CPI	0.72***	1.17***		
	(0.04)	(0.20)		
CPI-proxy			0.05	
			(0.06)	
CPI-replication				0.76
				(0.98)
CPI-proxy (t-1)			−0.15***	0.34***
			(0.03)	(0.09)
CPI-replication (t-1)			0.73***	−0.03
			(0.11)	(0.91)
Constant	0.71***	−1.38**	0.15	0.24
	(0.23)	(0.63)	(0.13)	(0.41)
N	57	57	56	56
R^2	0.86	0.53	0.60	0.14

Notes: Linear regressions of the replication and the proxy on the CPI (first and second columns). Regressions of the measurement errors of the replication and the proxy on current and lagged values of the measurement errors (third and fourth columns). The measurement errors are calculated as the difference between actual inflation and the error-ridden inflation rates. HAC-robust standard errors are given in parentheses. Coefficients with superscripts ***,**,* are statistically significant at the 1%, 5%, 10% level.

Table 3 provides regression results to show how the proxy and replication relate to actual inflation and whether the measurement errors are i.i.d. The first two columns regress the two error-ridden measures of inflation on actual CPI inflation, in line with the functional form assumed in the simulation exercise (see Eq. 5). For the replication, the slope is smaller than one and the constant positive. Therefore, the measurement errors are not of the classical type, which would require that the slope is one and the constant is zero. Meanwhile, the R^2 is relatively high (0.86) suggesting that the variance of the remaining measurement errors is low. For the proxy, the opposite pattern emerges. The slope is larger than one—although the difference is not statistically significant—, the constant significantly negative, and the R^2 is lower than for the replication. The third and fourth

columns show that the measurement errors are not i.i.d. The error of the replication is significantly related to its own lag and also related to past errors of the proxy.

Table 4 — Misclassification rates

	Replication		Proxy		Product		Joint	
	$d = 1$	$d = 0$	$d = 1$	$d = 0$	$d = 1$	$d = 0$	$d = 1$	$d = 0$
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

Notes: 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 gives the product of the misclassification rates for the replication and the proxy, which would correspond to the joint misclassification rate if the two indicators would be 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 gives the joint misclassification rates. The second and third row vary the threshold (average inflation instead of 0) and the sample period.

Table 4 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 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 shows results for a higher threshold set to the unconditional mean of CPI inflation.

For the entire sample, the replication never misclassifies deflation periods and hardly misclassifies inflation periods. The proxy is less accurate: it misclassifies 12% of inflation periods. The third panel in Table 4 shows 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 joint misclassification rates are close to zero and hardly differ.

These results are not suprising because inflation was relatively high during this sample period and the replication is less volatile than actual CPI inflation. Therefore, the second row shows 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 only to 11% (deflation) and 20% (inflation). This confirms our intuition that exploiting the information from two error-ridden indicators mitigates the misclassification bias. We also see, however, that under the conditional independence assumption, 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 U.S. industrial production from Davis (2004). His series is a quantity-based measure directly estimated in real terms. This measure is therefore not affected by measurement errors in price indices.[21] Second, I use various real activity measures for the U.K. because historical data for the U.K. is arguably of better quality than for other countries.[22] Third, I extend the analysis using GDP growth for eleven countries. Because measurement problems vary across countries these estimates are affected by the deflator bias to varying degrees.

4 The cost of deflation revisited

Table 5 provides estimates of average industrial production growth during periods with rising prices (α), the average shortfall during periods with falling prices (β), and the average growth with falling prices ($\alpha + \beta$). The first column uses the CPI by Officer and Williamson (2016). According to this classification, industrial production growth amounted to 6.5% during inflationary periods. During deflation, growth was not statistically significantly lower. Indeed, growth during 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 of economic growth and deflation.

The second to fourth columns provide bounds and point estimates with the CPI and the proxy, assuming the binary indicators are conditionally independent. The shortfall in industrial production growth during deflation becomes more pronounced. The point estimate suggests a statistically significant shortfall of -4.5 pp. Moreover, the bounds suggest the actual coefficient lies between -4.4 pp and -9.7 pp. The last column relaxes the conditional independence assumption assuming the joint misclassification rates equal 15%. The shortfall during deflation becomes even larger (-7.6 pp).

It is worth emphasizing the classification error does not bias the unconditional mean of industrial production growth. It only changes how we allocate growth to periods with rising and falling prices,

---

[21]Even U.S. 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. In between the annual real GDP figures rely heavily on the industrial production index by Davis (2004). Therefore, the deflator bias may mostly affect the benchmark years, but not the other periods.

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

Table 5 — U.S. industrial production growth during 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

Notes: Model: $y_t = \alpha + \beta x_t + \epsilon_t$ where $x_t \equiv \mathbf{1}_{\{\tilde{\pi}_t < 0\}}$. Model parameters: Mean growth rate during inflation α; shortfall during deflation β; mean growth rate during deflation $\alpha + \beta$; probability of deflation ($P\pi < 0$). Bias estimates: Bias if we would 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 a point estimates using the CPI and proxy, assuming conditional independence. Conditional dependence: Point estimates using the CPI and proxy, assuming the joint misclassification probabilities equal 0.15. HAC-robust standard errors are given in parentheses. ***,**,* denotes significance at the 1%, 5%, 10% level.

as well as, the share of deflation periods. Therefore, it is possible deflationary periods were, on average, still associated with robust growth. The mean growth rate during deflation declines when controlling for measurement errors. In no specification, however, is the growth rate significantly negative.

Are these biases statistically significant? Although we cannot directly compare the standard errors across models, GMM allows to estimate the bias as a non-linear function of the underlying coefficients.[23] The lower panel shows the bias estimates if we would only use the CPI. Under the conditional independence assumption, the shortfall in real activity is significantly biased by 1.4 pp. Allowing for conditional dependence we additionally find significant biases in the growth rate during inflation (-2.8 pp, 10% level), and the growth rate during 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 U.S. has the drawback that the country was little developed during the 19th century. Data is therefore necessarily scarce and the economy tilted towards agriculture. I therefore examine other countries as well. Using data for the U.K. has the advantage that various estimates of real activity exist. In addition, it is possible to use a second price index as a proxy.[24] Table 6 shows the change in various real activity measures during deflationary episodes. We find an increase in the unemployment rate, as well as declines in consumption, investment, and industrial production growth. In all specifications, the coefficients become larger in absolute size when exploiting a second deflation indicator.

The results suggest deflationary periods were associated with a relevant decline in economic activity. Why do these results differ from existing studies? First, I analyse a different sample period. But the results stay qualitatively similar when focusing on various subsamples, although the coefficients are less precisely estimated before 1870.[25] Second, other studies examined the link between inflation and real activity using GDP growth. But I find a significant decline in real U.S. GDP growth as well.[26] Finally, GDP data for other countries may be more affected by measurement problems in the GDP

---

[23]See Online Appendix A for a derivation.

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

[25]See Online Appendix C.

[26]See Online Appendix C.

Table 6 — Change in U.K. economic activity during 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

Notes: 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 during 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 the joint misclassification probabilities equal 0.15. HAC-robust standard errors are given in parentheses. ***,**,* denotes significance at the 1%, 5%, 10% level.

deflator. If this is the case, we expect to find a weaker, or even positive, link between deflation and real GDP growth for countries with more serious measurement problems. In what follows, I provide cross-country evidence in line with this interpretation.[27]

Figure 3 panel (a) shows the estimated OLS coefficients using GDP per capita growth for eleven countries.[28] A heterogeneous picture emerges. GDP growth is higher during deflation periods for Portugal, Finland, Switzerland, and Spain. Meanwhile, there is a negative association for the U.S., U.K., and Sweden. We also see that for countries with a high β, the estimate of α tends to be low.
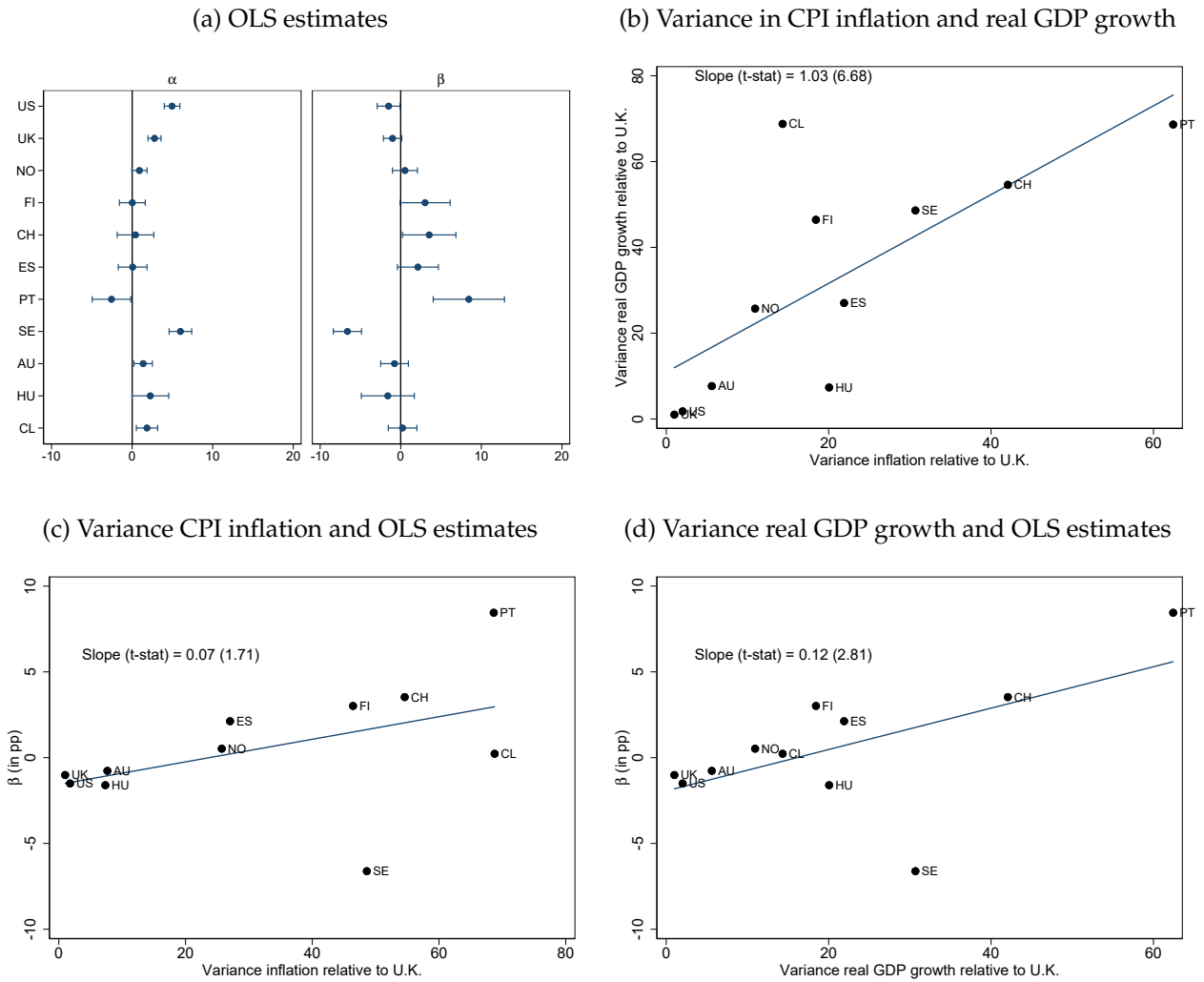
Against the backdrop of the simulations in Figure 1, this pattern may be caused by more serious measurement errors pushing the probability limit of the two coefficients in opposite directions. In other words, a mismeasured GDP deflator may be a more serious problem for Portugal and Switzerland than for the U.K. or the U.S.

Panel (b) shows the relationship between the volatility of inflation and real GDP growth. The variance is computed relative to the variance in the U.K., assuming that U.K. inflation and GDP growth are relatively well measured. If the same measurement errors appear in the CPI and the

---

[27]It is beyond the scope of the paper to judge the quality of the data for all these countries. At least for Switzerland there is evidence that 19th century CPI data are only rough estimates (see Ritzmann and David 2012). These estimates often use wholesale prices, cover a limited geographical area, and are very noisy (see Studer and Schuppli 2008; Kaufmann 2019).

[28]I am very grateful to an anonymous referee for providing these data.

Figure 3 — Volatility of historical data and OLS estimates

(a) OLS estimates

(b) Variance in CPI inflation and real GDP growth

(c) Variance CPI inflation and OLS estimates

(d) Variance real GDP growth and OLS estimates

Notes: Model: $y_t = \alpha + \beta x_t + \epsilon_t$ where $x_t \equiv \mathbf{1}_{\{\tilde{\pi}_t < 0\}}$. Panel (a): real GDP growth during inflation (α) and shortfall during deflation (β), with HAC-robust 95% confidence intervals. Panel (b): Relationship between the volatility of inflation and GDP growth. Panels (c)-(d): Relationship between the volatility of macroeconomic data and the OLS estimate of shortfall during deflation (β). All volatilities normalized by the volatility of U.K. data.

deflator, we would 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 would expect countries with more volatile measurement errors, and therefore more volatile inflation and GDP growth, to exhibit the largest distortion in the estimates. Panels (c) and (d) shows that there is a significant relationship between the shortfall during deflation (β) and the variance in GDP growth. Meanwhile, the correlation is only significant at the 10% level for the variance of CPI inflation. If we would exclude Sweden, which seems to be an outlier, the relationship would be clearly statistically significant.[29]

We can ask in addition whether the increases in GDP growth we find for some countries can be reproduced with reasonable assumptions on the volatility of the measurement errors. First, I construct an error-ridden GDP measure for the 19th century U.S., deflating nominal GDP with the CPI and the proxy.[30] This mimics nominal GDP is deflated by an error-ridden CPI based on wholesale prices (see e.g. Studer and 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 U.S. replications.[31]

The results are shown in Table 7. If we use the U.S. 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 real GDP growth in the U.S. relies heavily on industrial production by Davis (2004). If we use the proxy as a deflator we find a significantly positive association between deflation and real GDP growth, at least when exploiting multiple indicators.[32]

The simulation exercise confirms these results qualitatively. 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. But,

---

[29]See Online Appendix C.

[30]See Online Appendix B, Figure B.2.

[31]I assume $\alpha = 3$ and $\beta = -1$ and set the standard deviation of inflation to 2.8 (see Table 2). I then construct the deflation dummy 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 2. 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.

[32]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 it was possible to assemble a second price measure. The relationship between the OLS coefficients and the volatility of real GDP growth and inflation becomes more pronounced.

Table 7 — The role of the GDP deflator

	Baseline	Cond. independence			Cond. dependence
U.S. 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)
U.S. 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

Notes: Model: $y_t = \alpha + \beta x_t + \epsilon_t$ where $x_t \equiv \mathbf{1}_{\{\tilde{\pi}_t < 0\}}$. The first two specifications use U.S. GDP growth deflated by either the CPI or the proxy. The estimation sample ranges from 1800–1899. The remaining specifications use 1,000 simulated observations under different assumptions on the measurement errors. Model parameters: Mean growth rate during inflation α; shortfall during 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 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.

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.

4.1 Robustness

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

The literature suggests asset price declines matter more than CPI deflation (see e.g. Borio et al. 2015). In addition, frequent banking crises may cause deflation. Jalil (2015) shows that the U.S. price level declined significantly after major banking panics. Banking crises may also be a consequence of tight monetary policy and deflation (see e.g. Friedman and Schwartz 1963, p. 108). I therefore add dummies for stock price declines, banking crises, and money slowdowns.[33] Controlling for additional covariates leaves the association between deflation and real activity intact. The bounds, as well as the GMM point estimates, suggest deflation is associated with significantly lower industrial production growth.

The results are robust when taking into account more than two deflation indicators. I obtained an alternative price index by Falkner from 1840–1891 (Aldrich Report 1893, p. 93).[34] This price index is not included in the composite CPI by Officer and Williamson (2016) because of its many deficiencies.[35] 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 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 using the Falkner index instead of our proxy yields qualitatively similar results.

I also varied the definition of the deflation classification to take into account some deflationary periods may be more harmful than others (see Bordo and 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 industrial production growth declines at least by 1 pp more. The results are also robust when focusing on persistent deflations of two or more years. The OLS estimate amounts to about -1.6 pp and is not

---

[33]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.

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

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

statistically significant. After 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 our assumptions for the joint misclassification rates I estimate the model for all combinations of the joint probability of misclassifying inflation and deflation, respectively, between 0% and 15% in steps of 5%. For most combinations, we find a statistically significant shortfall during deflation. In addition, the bias is usually statistically significantly different from zero.

The estimators assume that the error term is i.i.d. If this assumption is violated the GMM estimate is biased (see Cosslett and Lee 1985). Black et al. (2000) suggest that the bounding strategy still works if the measurement errors are uncorrelated with other covariates. I therefore provide OLS estimates controlling for one lag of the dependent variable. The results do not change because U.S. 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 deflations were benign, short-lived, or a consequence of beneficial advances in productivity. This paper proposes an additional explanation: measurement problems in historical price data.

Is deflation costly after all? This paper suggests deflation is at least more costly than we thought. Although the results cannot directly be extrapolated to the present, the findings have implications for today: 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 an increasing number of studies using historical cross-country data (see e.g. Reinhart and Rogoff 2010; Schularick and Taylor 2012; Jordá et al. 2016; Baker et al. 2018). Importantly, adding an additional country may distort the results if the data quality is poor. In addition, if measurement problems are common across multiple series, the biases are difficult to anticipate without detailed knowledge of the underlying data. In fact, these biases may well offset the gains in terms of more efficient estimates.

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), would be an interesting avenue for future research.

References

Adams, Thurston 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," Vermont Agricultural Experiment Station, University of Vermont and State Agricultural College.

Aghion, Philippe, Antonin Bergeaud, Timo Boppart, Peter J. Klenow, and Huiyu Li (2017) "Missing Growth from Creative Destruction," Working Paper Series 2017-4, Federal Reserve Bank of San Francisco.

Aigner, Dennis J. (1973) "Regression With a Binary Independent Variable Subject to Errors of Observation," *Journal of Econometrics*, Vol. 1, pp. 49–60.

Aldrich Report (1893) "Wholesale Prices, Wages, and Transportation," Senate Committee on Finance, 52nd Congress, 2nd Session, Report 1394, Part 2.

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

Atkeson, Andrew and Patrick J. Kehoe (2004) "Deflation and Depression: Is There an Empirical Link?," *American Economic Review*, Vol. 94, pp. 99–103.

Baker, Sarah S., David López-Salido, and Edward Nelson (2018) "The Money View Versus the Credit View," Discussion Paper 12982, C.E.P.R.

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

Bayoumi, Tamim and Barry Eichengreen (1996) "The Stability of the Gold Standard and the Evolution of the International Monetary Fund," in Tamim Bayoumi, Barry Eichengreen, and Mark P. Taylor eds. *Modern Perspectives on the Gold Standard*, Cambridge: Cambridge University Press, Chap. 6, pp. 165–188.

Beckworth, David (2007) "The Postbellum Deflation and its Lessons for Today," *The North American Journal of Economics and Finance*, Vol. 18, pp. 195–214.

Benati, Luca (2008) "Investigating Inflation Persistence Across Monetary Regimes," *The Quarterly Journal of Economics*, Vol. 123, pp. 1005–1060.

Bernanke, Ben S. (2010) "The Economic Outlook and Monetary Policy," Speech at the Federal Reserve Bank of Kansas City Economic Symposium, Jackson Hole, Wyoming, `www.federalreserve.gov/newsevents/speech/bernanke20100827a.htm`, accessed on 2 November 2019.

Black, Dan A., Mark C. Berger, and Frank A. Scott (2000) "Bounding Parameter Estimates with Nonclassical Measurement Error," *Journal of the American Statistical Association*, Vol. 95, pp. 739–748.

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

Bordo, Michael D. and Angela Redish (2004) "Is Deflation Depressing? Evidence From the Classical Gold Standard," in Richard C.K. Burdekin and Pierre L. Siklos eds. *Deflation: Current and Historical Perspectives*, Cambridge: Cambridge University Press, Chap. 7, pp. 99–202.

Borio, Claudio, Magdalena Erdem, Andrew Filardo, and Boris 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," `www.ssa.gov/history/reports/boskinrpt.html`, accessed on 31 october 2019.

Bound, John and Alan B. Krueger (1991) "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics*, Vol. 9, pp. 1–24.

Card, David (1996) "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis," *Econometrica*, Vol. 64, pp. 957–979.

Cogley, Timothy and Thomas J. Sargent (2015) "Measuring Price-Level Uncertainty and Instability in the United States, 1850-2012," *The Review of Economics and Statistics*, Vol. 97, pp. 827–838.

Cosslett, Stephen R. and Lung-Fei Lee (1985) "Serial Correlation in Latent Discrete Variable Models," *Journal of Econometrics*, Vol. 27, pp. 79–97.

David, Paul A. and Peter Solar (1977) *A Bicentenary Contribution to the History of the Cost of Living in America*, Vol. 2 of Research in Economic History, Greenwich: JAI Press.

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

Eichengreen, Barry, Donghyun Park, and Kwanho Shin (2016) "Deflation in Asia: Should the Dangers be Dismissed?," Asian Development Bank Economics Working Paper Series 490.

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

Goolsbee, Austan D. and Peter J. Klenow (2018) "Internet Rising, Prices Falling: Measuring Inflation in a World of E-Commerce," Working Paper 24649, National Bureau of Economic Research.

Griliches, Zvi (1986) "Economic Data Issues," in Z. Griliches and M. D. Intriligator eds. *Handbook of Econometrics*, Vol. 3: Elsevier, Chap. 25, pp. 1465–1514.

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

Hartmann, Philipp and Frank Smets (2018) "The First Twenty Years of the European Central Bank: Monetary Policy," Working Paper Series 2219, European Central Bank.

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

Hoover, Ethel D. (1958) "Wholesale and Retail Prices in the Nineteenth Century," *The Journal of Economic History*, Vol. 18, pp. 298–316.

———— (1960) "Retail Prices after 1850," in *Trends in the American Economy in the Nineteenth Century*: Princeton University Press, pp. 141–190.

HSSO (2012) "Historische Statistik der Schweiz HSSO," Tab. Q.1a, www.hsso.ch/2012/q/, accessed on 22 October 2019.

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

Johnston, Louis and Samuel H. Williamson (2019) "Sources and Techniques Used in the Construction of Annual GDP, 1790-1928," www.measuringworth.com/usgdp/, accessed on 22 october 2019, MeasuringWorth.

Óscar Jordá, Moritz Schularick, and Alan M. Taylor (2016) "Macrofinancial History and the New Business Cycle Facts," in *NBER Macroeconomics Annual 2016*, Vol. 31: University of Chicago Press.

Kane, Thomas J., Cecilia Elena Rouse, and Douglas Staiger (1999) "Estimating Returns to Schooling When Schooling is Misreported," NBER Working Papers 7235, National Bureau of Economic Research.

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

Kreider, Brent (2010) "Regression Coefficient Identification Decay in the Presence of Infrequent Classification Errors," *The Review of Economics and Statistics*, Vol. 92, pp. 1017–1023.

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

Long, Clarence D. (1960) *Wages and Earnings in the United States, 1860-1890*: Princeton University Press.

Mahajan, Aprajit (2006) "Identification and Estimation of Regression Models with Misclassification," *Econometrica*, Vol. 74, pp. 631–665.

Margo, Robert A. (2000) *"Wages and Labor Markets in the United States, 1820-1860*: University of Chicago Press.

Officer, Lawrence H. (2014) "What Was the Consumer Price Index Then? A Data Study," www.measuringworth.com/docs/cpistudyrev.pdf, accessed on 31 october 2019, MeasuringWorth.

Officer, Lawrence H. and Samuel H. Williamson (2016) "Annual Consumer Price Index for the United States, 1774-2015," www.measuringworth.com/uscpi/, accessed on 22 october 2019, MeasuringWorth.

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

Reinhart, Carmen M. and Kenneth S. Rogoff (2010) "Growth in a Time of Debt," *American Economic Review*, Vol. 100, pp. 573–578.

Ritzmann, Heiner and Thomas David (2012) "Schätzung des Bruttoinlandprodukts nach Branchen und Kantonen 1890-1960," in Patrick Halbeisen, Margrit Müller, and Béatrice Veyrassat eds. *Wirtschaftsgeschichte der Schweiz im 20. Jahrhundert*, Basel: Schwabe, Chap. A.2.

Romer, Christina (1986a) "Is the Stabilization of the Postwar Economy a Figment of the Data?," *American Economic Review*, Vol. 76, pp. 314–3347.

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

Schularick, Moritz and Alan M. Taylor (2012) "Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870-2008," *American Economic Review*, Vol. 102, pp. 1029–1061.

Shoemaker, Owen J. (2014) "Variance Estimates for Price Changes in the Consumer Price Index: January-December 2014," www.bls.gov/cpi/tables/variance-estimates/home.htm, accessed on 31 october 2019, U.S. Bureau of Labour Statistics.

SNB (2011) "Swiss National Bank Sets Minimum Exchange Rate at CHF 1.20 Per Euro," https://www.snb.ch/en/mmr/reference/pre_20110906/source/pre_20110906.en.pdf, accessed on on 2 november, 2019, Swiss National Bank.

Studer, Roman and Pascal Schuppli (2008) "Deflating Swiss Prices over the Past Five Centuries," *Historical Methods*, Vol. 41, pp. 137–153.

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

Weeks, Joseph D. (1886) "Report on the Statistics of Wages in Manufacturing Industries," Department of the Interior, Census Office.