FEDERAL RESERVE BANK OF ST. LOUIS

REVIEW

FOURTH QUARTER 2023
VOLUME 105 | NUMBER 4

REVIEW

Volume 105 • Number 4

FEDERAL RESERVE BANK *of* ST. LOUIS
CENTRAL TO AMERICA'S ECONOMY®

Review
Review is published four times per year by the Research Division of the Federal Reserve Bank of St. Louis. Full online access is available to all, free of charge.

Online Access to Current and Past Issues
The current issue and past issues dating back to 1967 may be accessed through our Research Division website: http://research.stlouisfed.org/publications/review. All nonproprietary and nonconfidential data and programs for the articles written by Federal Reserve Bank of St. Louis staff and published in *Review* also are available to our readers on this website.

Review articles published before 1967 may be accessed through our digital archive, FRASER: http://fraser.stlouisfed.org/publication/?pid=820.

Review is indexed in Fed in Print, the catalog of Federal Reserve publications (http://www.fedinprint.org/), and in IDEAS/RePEc, the free online bibliography hosted by the Research Division (http://ideas.repec.org/).

Subscriptions and Alerts
The *Review* is no longer printed or mailed to subscribers. Our last printed issue was the first quarter of 2020.

Our monthly email newsletter keeps you informed when new issues of *Review*, *Economic Synopses*, *Regional Economist*, and other publications become available; it also alerts you to new or enhanced data and information services provided by the St. Louis Fed. Subscribe to the newsletter here: http://research.stlouisfed.org/newsletter-subscribe.html.

Economic Data
General economic data can be obtained through FRED®, our free database with over 800,000 national, international, and regional data series, including data for our own Eighth Federal Reserve District. You may access FRED through our website: https://fred.stlouisfed.org.

ISSN 0014-9187

FEDERAL RESERVE BANK of ST. LOUIS
CENTRAL TO AMERICA'S ECONOMY®

Fiscal Dominance and the Return of Zero-Interest Bank Reserve Requirements

Charles W. Calomiris

Abstract

As a matter of arithmetic, the trends of US government debt and deficits will eventually result in an outrageously high government debt-to-GDP ratio. But when exactly will the United States hit the constraint of infeasibility and how exactly will policy adjust to it? This article considers fiscal dominance, which is the possibility that accumulating government debt and deficits can produce increases in inflation that "dominate" central bank intentions to keep inflation low. Is it a serious possibility for the United States in the near future? And how might various policies change (especially those related to the banking system) if fiscal dominance became a reality?

JEL codes: E52, E62, H63

Federal Reserve Bank of St. Louis *Review*, Fourth Quarter 2023, *105*(4), pp. 223-33.
https://doi.org/10.20955/r.105.223-33

> Under current policy and based on this report's assumptions, [government debt relative to GDP] is projected to reach 566 percent by 2097. The projected continuous rise of the debt-to-GDP ratio indicates that current policy is unsustainable.
>
> —*Financial Report of the United States Government*, February 16, 2023

INTRODUCTION

The above quotation from the Treasury's Financial Report admits that the current combination of government debt and projected deficits is not feasible as a matter of arithmetic because it would result in an outrageously high government debt-to-GDP ratio. But when exactly will the US hit the constraint of infeasibility, and how exactly will US policy adjust to it? This article considers whether fiscal dominance is a serious possibility for the United States in the near future and discusses how various policies (especially those related to the banking system) likely would change if fiscal dominance became a reality.

Fiscal dominance refers to the possibility that the accumulation of government debt and continuing government deficits can produce increases in inflation that "dominate" central bank intentions to keep

Charles W. Calomiris is Henry Kaufman Professor of Financial Institutions Emeritus at Columbia Business School, and Dean of the Center for Economics, Politics and History at UATX. The author thanks Peter Ireland, Peter Stella, the referee, the editor, and participants in the Alamos Alliance 2023 meetings for helpful comments on earlier drafts.

inflation low. The article begins by showing that the prospect of this occurring soon in the United States is no longer far-fetched. Indeed, if global real interest rates returned tomorrow to their historical average of roughly 2 percent, given the existing level of US government debt and large continuing projected deficits, the US would likely experience an immediate fiscal dominance problem. Even if interest rates remain substantially below their historical average, if projected deficits occur as predicted, there is a significant possibility of a fiscal dominance problem within the next decade.

The essence of fiscal dominance is the need for the government to fund its deficits on the margin with non-interest-bearing debts. The use of non-interest-bearing debt as a means of funding is also known as "inflation taxation." Fiscal dominance leads governments to rely on inflation taxation by "printing money" (increasing the supply of non-interest-bearing government debt).[1] To be specific, here is how I imagine this occurring: When the bond market begins to believe that government interest-bearing debt is beyond the ceiling of feasibility, the government's next bond auction "fails" in the sense that the interest rate required by the market on the new bond offering is so high that the government withdraws the offering and turns to money printing as its alternative.

As the money supply is forced to grow by fiscal dominance, inflation rises, which creates a new means of funding government expenditures via "inflation taxation." Inflation taxation has two components: expected and unexpected inflation taxation. Both are limited in their ability to fund *real* government expenditures. The expected component of inflation taxation (per period) is the product of the nominal interest rate and the inflation tax base, which consists of all non-interest bearing government debt. (Typically, this consists of currency and non-interest-bearing bank reserves at the central bank.) Total real government expenditures that can be financed by the expected inflation tax are limited because the tax base of this inflation tax is determined by the demand for money. The inflation tax earned per period is the product of the nominal interest rate (the inflation tax rate) and the amount of real demand for currency and zero-interest reserves. Unexpected inflation taxation occurs when the nominal value of outstanding government debt falls unexpectedly (thereby taxing government debtholders), and this component is also limited by the ability of government to surprise markets by creating unanticipated inflation.

If the US government faced a fiscal dominance problem, it would have to fund real deficits by real inflation taxation, which is a limited tax resource. Thus, not all real deficits are feasible to fund with inflation taxation.

Furthermore, some changes in policy with respect to reserve requirements are likely if fiscal dominance becomes a reality. The existing amount of the zero-interest debt (the inflation tax base) is currently limited to only currency, given that bank reserves bear interest today. Given the small size of the currency outstanding, if the government wishes to fund large real deficits, that will be easier to do if the government eliminates the payment of interest on reserves. This potential policy change implies a major shock to the profits of the banking system.

Second, as the history of inflation episodes has shown, even an inflation tax base of currency plus zero-interest reserves would decline in real terms in the face of a significant increase in inflation. Based on data for the US as of 2023, the resulting inflation rate could be very high. (That rate is derived by calculating the inflation rate that, when multiplied by the inflation tax base, results in inflation taxation sufficient to fund projected deficits.)

For that reason, it is quite possible that a fiscal dominance episode in the US would result in not only the end of the policy of paying interest on reserves, but also a return to requiring banks to hold a large fraction of their deposit liabilities as zero-interest reserves. For example, under one illustrative example, I will show that requiring banks to hold 40 percent of deposits as zero-interest reserves, under

1. For an example of how fiscal dominance can produce inflation taxation in the steady state, see the study of Brazilian inflation by Calomiris and Domowitz (1989).

reasonable assumptions, would reduce the annual inflation rate to fund likely deficits from an inflation of about 16 percent to only about 8 percent. For that reason, imposing high reserve requirements for zero-interest paying reserves may seem quite attractive to a policymaker interested in reducing the inflationary consequences of fiscal dominance.

The history of inflation taxation around the world has shown that when governments become strapped for resources, they often use zero-interest reserve requirements to tax banking systems and remove their spending constraints. For example, in Mexico during the 1970s and early 1980s, inflation taxation of banks became increasingly relied on as government expenditures rose; eventually, as fiscal problems mounted, the government expropriated first bank depositors and then bank equity holders by nationalizing the banks (Calomiris and Haber, 2014, Chapter 11). The general problem of impecunious governments taxing banks with the inflation tax, credit controls, or other means—which can have major adverse consequences for efficient capital allocation and growth—is the theme of a very large literature, which goes back at least as far as Gurley and Shaw (1960) and includes such landmark contributions as McKinnon (1973), Fry (1988), and Acharya (2020).

Taxing banks with reserve requirements and zero-interest reserves is convenient for two reasons. First, instead of new taxes enacted by legislation (which may be blocked in the legislature), reserve requirements are a regulatory decision that is generally determined by financial regulators. It can be implemented quickly, assuming that the regulator with the power to change the policy is subject to pressure from fiscal policy. In the case of the US, it is the decision of the Federal Reserve Board whether to require reserves to be held against deposits and whether to pay interest on them.

Second, because many people are unfamiliar with the concept of the inflation tax (especially in a society that has not lived under high inflation), they are not aware that they are actually paying it, which makes it very popular among politicians. If, as I argue below, a policy that would eliminate interest on reserves and require a substantial proportion of deposits to be held as reserves would substantially reduce inflation, then I believe it would be hard for the Federal Reserve Board to resist going along with that policy.

Such a policy change would not only reduce bank profitability but also reduce the real return earned on bank deposits to substantially below other rates of return on liquid assets, which potentially could spur a new era of "financial disintermediation," as consumers and firms seek alternatives to low-interest paying bank deposits. Such financial disintermediation from banks occurred in the US in the 1960s and 1970s as a result of high inflation and regulations (including both zero-interest reserve requirements and ceilings on deposit interest rates) that limited the interest rates banks could pay to depositors. Of course, banks and their political allies might try to oppose financial innovations to allow firms and consumers to exit from banks, which would lead to a potentially interesting regulatory battle over the future of financial intermediation. The need to preserve a high inflation tax base could lead to a political choice to preserve a technologically backward banking system. (This would be a continuation and acceleration of recent political trends to limit Fintech bank chartering, as discussed in Calomiris, 2021.)

An alternative policy path, of course, with less inflation taxation, would be for the government to decide to reduce fiscal deficits and thereby avoid the need for rising inflation and its adverse consequences for the banking system. This may be a hard policy to enact, however, given that the main contributors to future deficits are large Medicare and Social Security entitlement payments. Also, defense spending seems likely to rise as the result of increasing geopolitical risks related to China. Increased income taxation is another alternative, but this too may be unlikely, not only because of the lack of political consensus about taxation but also because it would reduce growth in income, which would partly offset any deficit reduction coming from projected increases in the ratio of taxes to income. Ultimately, it seems likely that the US will either have to decide to rein in entitlements or risk a future of significantly higher inflation and financial backwardness.

WHEN DOES FISCAL DOMINANCE BECOME APPARENT AS A CONSTRAINT?

Historically, high inflation is produced by growth in the supply of money that reflects the pressures of fiscal dominance. Every major inflation in world history is a fiscal phenomenon before it is a monetary phenomenon.

When exactly does fiscal dominance lead to monetization, and how much does that threaten monetary policy's inflation goals? How much dollar inflation might result from excessive US government borrowing?

If a country issues debt denominated only in its own currency (which is true of the US), then fiscal dominance arithmetic requires one to focus on only one equation ("transversality condition"), the violation of which forces inflation to rise. Here, the condition (formalized in McCallum, 1984) characterizes how the total debt held by the public is limited by the present value of prospective future fiscal surpluses. At some point, as the government debt-to-GDP ratio rises sufficiently (for a given real interest rate and rate of real growth of the economy), the real demand for government debt reaches a maximum. Any issues of nominal debt beyond that amount will not be accepted by the bond market as increases in real debt.

Where is the ceiling? If nominal GDP is $24 trillion, and current interest-bearing debt not held by government agencies or by the Fed is about $20 trillion, how much higher can the debt-to-GDP ratio go, given our current long-term real interest rate and real rate of growth?[2] First, note that much of the Fed-held debt nowadays pays interest (whether through $3 trillion in excess reserves to Fed member banks or $2.4 trillion in Fed reverse repos mainly with money market mutual funds). If the Fed is paying interest on the debt, then this should be added to the Treasury debt held by the public. So that means that the current size of interest-bearing public debt is greater than GDP. This is a historical high in the debt-to-GDP ratio reached only once before, during World War II. Not only that, but the deficit has been running, and is projected to continuing running, at greater than 5 percent of GDP per year. CBO projections imply that the debt-to-GDP ratio will reach about 200 percent by 2050, but that assumes that real interest rates will remain at their historical lows; if they rise even by a small amount, the implied debt growth scenario is much worse.

Is an interest-bearing-debt-to-GDP ratio of greater than 100 percent, combined with current and projected deficits of that magnitude, an immediate danger for inflation? The answer is sensitive to your beliefs about future government spending and taxes, about the government's ability to change spending and taxes to avoid issuing new debt if that becomes a problem, and about the path of future real interest rates and economic growth. The current debt-to-GDP ratio is in the range associated historically with the fiscal dominance ceiling. On the other hand, real interest rates are at historical lows, which means that higher debt-to-GDP ratios are feasible. It is hard to say precisely where the fiscal dominance ceiling is because that requires one to forecast economic growth, taxes, government expenditures, and real interest rates.

As a country gets close to the ceiling, the interest rates creditors demand rise (as they demand an inflation risk premium) in anticipation of the risk of hitting the ceiling. That means that the perceived risk of hitting the ceiling (once you are close to it) can become self-fulfilling.

Rather than focus on the ceiling itself, it's better to ask what is the highest ratio of debt-to-GDP that is safely distant from the ceiling so that the government does not stumble into a self-fulfilling fiscal dominance equilibrium. I don't know the answer of where danger will arise, but there is a clear risk now that a rise in global real interest rates could reverse their three decades of decline (which remains poorly understood), which would make even the current debt-to-GDP ratio very risky, especially given that

2. Government debt held by other government agencies is excluded from the computations below.

large primary deficits are projected to continue and that spending cuts are hard because so much of government spending in the US is from entitlements (not subject to annual appropriations) or from military spending (which is unlikely to be reduced, given the current geopolitical landscape). Taxes could be increased, but it seems unlikely to think that a divided Congress today would be nimble enough to do that to stop an incipient inflation risk problem from emerging.

What does a government that goes above the ceiling, or even gets too close to the ceiling, do to survive? It "prints money," which you can imagine as using cash to pay for government bills rather than issuing new interest-bearing debt. Mechanically, in the US, that would be accomplished by Fed open market purchases of government debt (so-called deficit monetization). When the Fed buys government debt, the reason that is a useful thing to do is that it reduces the amount of outstanding interest-bearing government debt in the hands of the public. Interest-bearing government debt pays an interest rate that rises with inflation (one for one). But cash (or reserves that pay zero interest) pays no interest. Monetization allows the Treasury to sell debt to the Fed that public bondholders are unwilling to finance.

Before explaining the policy choices the Fed would face in monetizing deficits under fiscal dominance, it is worth noting that, for many other countries, the debt-to-GDP ceiling math is not the only fiscal math to be considered. If a country has issued some of its debt in foreign currency, things become more complicated because foreign debt can be at risk of default even if the first fiscal dominance transversality condition (based on total debt, GDP growth, and the real interest rate) is not violated. A second transversality condition connects the outstanding amount of *externally denominated debt* to prospective future net exports. An increased risk of foreign-denominated debt default due to the risk of violating the second transversality condition may lead to an increase in inflation, too, through an indirect channel: inflation through monetization of debt can be useful as a means of avoiding bond issues that might increase the risk of default on foreign-denominated debts. There is some evidence related to this possibility. Accelerations of inflation tend to precede deteriorations in foreign debt ratings, which may be interpreted as evidence that raising domestic inflation is used to reduce the effects of rising fiscal deficits on foreign debt default (Cantor and Packer, 1996).

For countries like the US, however, which issues debt only in its home currency, the first transversality condition is the only one that matters. For other countries (especially emerging market countries), both conditions are important, and the second transversality condition can be violated even when the first is not at risk (as many believed was the case in Argentina in 2001).

LIMITS OF INFLATION TAXATION

If the US were forced by debt-to-GDP math to fund itself by "printing money," how would it do so and what are the limits to how much real government spending can be funded in this way? The real funding of spending from the "inflation tax" (the funding of its expenditures by printing money) is limited. There are two components to the inflation tax: expected inflation taxation and unexpected inflation taxation. The expected component of the inflation tax is limited by the real demand for zero-interest government money. The unexpected component is limited by the one-time gain that comes from surprising people with rising inflation, which reduces the value of existing government debt and the debt-to-GDP ratio, thereby creating new debt capacity for additional future deficits.

The real *expected* component of the inflation tax (the amount of real goods and services that can be paid for by printing money) has an upper bound because real demand for zero-interest money declines as inflation increases (see, for example, Sargent, 1982; Calomiris and Domowitz, 1989). That means that once a government exhausts its ability to fund with the inflation tax, it must reduce spending.

The government also may get another *unexpected* inflation tax related to the decline in the value of its outstanding bonds at the time the fiscal dominance problem arises. When fiscal dominance hits and leads to monetization, if this is not anticipated sufficiently far in advance, it also causes some or all existing bonds (long-term bonds with existing low coupons that aren't indexed to inflation) to fall in nominal value. This is a one-time gain to the government because, going forward, the government will pay a market interest rate on all new debt issues that incorporates the future rate of inflation. If the average duration of government debt is sufficiently long, and fiscal dominance is not anticipated years in advance, the government could benefit from a substantial capital gain from the unexpected inflation tax, which increases its real capacity to issue new interest-bearing debt by a similar amount.

This assumes that the public does not see the fiscal dominance problem coming; but in today's world, that does not seem as far-fetched an assumption as it used to be. It is also worth noting that the inflation surprise of the 1960s and 1970s entailed a very large capital gain from the unexpected inflation tax, as Robert Eisner documents in his book, *How Real Is the Federal Deficit?* Still, the ability to surprise the market with an acceleration of inflation is always limited by the fact that market participants monitor political and economic news closely and have knowledge about the processes that give rise to inflation.

It is also worth noting, as Beckworth (2023) points out, that in recent months the *market* value of US Treasury securities has already taken a toll on bondholders (including banks), declining from 108 percent of GDP in 2020 to about 85 percent of GDP today, mainly as the result of inflation surprises since 2021. This unexpected inflation tax has already substantially expanded the ability of the government to issue debt. More surprises may be coming, although the ability to surprise savers will decline going forward: As the saying goes, you can't fool all the people all the time.

What does printing money mean in practical terms? For the US today, the "tax base" for the expected component of the inflation tax does not include the $3 trillion dollars of reserves held by banks—because those pay interest. Of course, the Federal Reserve could change that policy (and perhaps would have to do so in the event of a fiscal dominance problem); but if it did not do so, the inflation tax would be earned only on the dollars in cash held by the public worldwide (about $2.2 trillion in nominal terms today).

The big questions about fiscal dominance problems for the US, therefore, are (1) how much would the government gain from surprising outstanding bondholders with fiscal dominance, (2) what would be the tax base of the expected inflation tax, and (3) how much would real demand for zero-interest cash fall (thereby reducing the real tax base of the expected inflation tax) as the result of an increase in expected inflation? Answering these questions is important because it tells us how much inflation would be needed for the government to be able to fund the fiscal costs not funded by interest-bearing debt (as a result of hitting the fiscal dominance threshold). Indeed, it is possible that when answering these questions one might even conclude that it is not feasible for the projected real fiscal deficits to be funded by the inflation tax. In that case, some cuts to governments spending or increases in other taxation would be necessary.

CONNECTING THE SIZE OF DEBT INCREASES, MONEY INCREASES, AND INFLATION

Any attempt to estimate how a fiscal dominance scenario would play out in the US is fraught with uncertainty. What follows is an illustration using an example.

Assume that the US suddenly reaches the maximum for total real government debt demand by the public. What would it do, how much would the price level have to rise, and how limited would its ability to fund spending with inflation be? To be concrete: If deficits required a one-time increase in money printing by, say, roughly 5 percent of outstanding government debts to pay existing bills, how much would inflation rise?

To make things clear in a simplified example, assume that at the time the fiscal dominance shock hits, inflation is zero, the real interest rate is zero, and real growth in the future is zero. Assume (using current data) that government debt held by the public, not including the Fed (G) is 20, that repo repurchases by the Fed (R) is 2.4, that reserves held by banks as interest-bearing excess reserves (M) is 3, and that cash held by the public is 2.2. I will assume that the annual real deficit that must be funded by the inflation tax is 3 percent of $24 trillion in 2023 dollars (which is realistic, given that the deficit is over 5 percent of GDP and real growth is roughly 2 percent).

I will assume that fiscal dominance will force the Fed to stop paying interest on reserves, but it must continue to pay interest on reverse repos, so the tax base for the expected component of the inflation tax is (C+M).[3]

The real demand for currency and the real demand for bank reserves at the Fed both should fall when increases in their nominal supplies cause prices to rise. In my example, I will assume realistically that if there is no reserve requirement, then the real demand falls more for any increase in inflation because both the public (depositors) and the banks (reserve holders) have incentives and ability to economize on real balances when faced with inflation. But when reserve requirements are set as a fraction of deposits, if they are a binding constraint, then only depositors can economize in reaction to inflation. So, real demand for reserves falls by less than it would otherwise. I also assume that the real demand for cash also falls in reaction to inflation. Judging from prior statistical studies, all these reactions that reduce the real inflation tax base (at relatively low rates of inflation) will reduce the real tax base more slowly than the rise in inflation will raise the inflation tax rate.[4] So a rise in inflation will increase the inflation tax, albeit not as quickly as the rise in the rate of inflation itself.

To make all the above assumptions concrete, assume that, in the *absence* of a reserve requirement, the inflation rate (p) solves the following equation (where (1 – p)3 is real reserve demand) and (1 – p/2)2.2 is real currency demand. In this simple formulation, if inflation rises from 0 to 10 percent, then real reserves demand falls by 10 percent and real currency demand falls by 5 percent. These specific "guesses" are realistic, and useful for our example, but of course, they are just guesses. One cannot be sure how real demand for currency or reserves will respond to changes in policy regimes. To solve for the inflation rate that generates sufficient inflation tax revenue to pay the real deficit per year, one solves

$$p(1 - p)3 + p(1 - p/2)2.2 = 0.73. \tag{1}$$

Recall that total reserves in 2023 dollars is $3 trillion, and currency held by the public is $2.2 trillion. The total deficit that needs funding is $0.73 trillion. The coefficient (1 – p/2) in the second expression captures the rate at which real demand for currency falls as inflation rises. The coefficient (1 – p) in the first expression captures the assumption that real demand for reserves can fall relatively quickly (by the rate (1 – p)) when there is no reserve requirement. The implied rate of inflation that satisfies the above equation is about 16.3 percent.

If a zero-interest reserve requirement of 40 percent is imposed on all deposits, and if that is a binding constraint, then reserve holdings rise to 7.2 (which is 40 percent of $18 trillion in commercial bank deposits observed in February 2023). Now real reserve demand is assumed to fall more slowly in reaction to inflation than it did without the imposition of the reserve requirement (at the rate 1 – p/2). Equation 1 can be rewritten as

$$p(1 - p/2)7.2 + p(1 - p/2)2.2 = 0.73. \tag{2}$$

3. It is also possible that the Fed could reduce but not eliminate interest payments on reserves. In my view, given the high inflation rate implied by fiscal dominance, the Fed would choose to eliminate all interest on reserves.

4. Cagan's (1956) formulation assumed that money demand took semi log form. The semi-elasticity of demand is alpha in that model, and the revenue-maximizing rate of inflation is 1/alpha.

The implied rate of inflation that satisfies this equation is about 8.1 percent. This implies a large reduction from inflation that is possible by raising the inflation tax base by mandating that banks hold zero-interest reserves as a large proportion of deposits.

At the opposite extreme of policy with respect to the inflation tax base, what if in the face of a fiscal dominance shocks, the Fed did not eliminate interest on reserves and paid the market interest rate on all reserves? Then the inflation tax would be charged entirely on the inflation tax base of the real demand for currency, which would mean

$$p(1 - p/2)2.2 = 0.73. \tag{3}$$

Under this assumed demand function, the result would be a rate of inflation in excess of 41 percent. Clearly, given that prospective outcome, the Fed will be under substantial pressure in a fiscal dominance situation to eliminate interest on reserves and impose a large reserve requirement.

This calculation likely is not correct, however: Past experience suggests that, at high rates of inflation, the real demand for currency falls at a faster rate than the rate of inflation rises and the revenue-maximizing rate of inflation is typically in the range of 30 to 45 percent (Cagan, 1956; Kimbrough, 2006). For example, if currency demand is $(1 - p)2.2$, then p should solve

$$p(1 - p)2.2 = 0.73. \tag{4}$$

But there is no rate of inflation, p, that satisfies Equation 4 because the maximum inflation tax that is feasible (assuming this real demand) is $0.55 < 0.73$ and the maximum revenue from the inflation tax occurs when p = 50 percent.

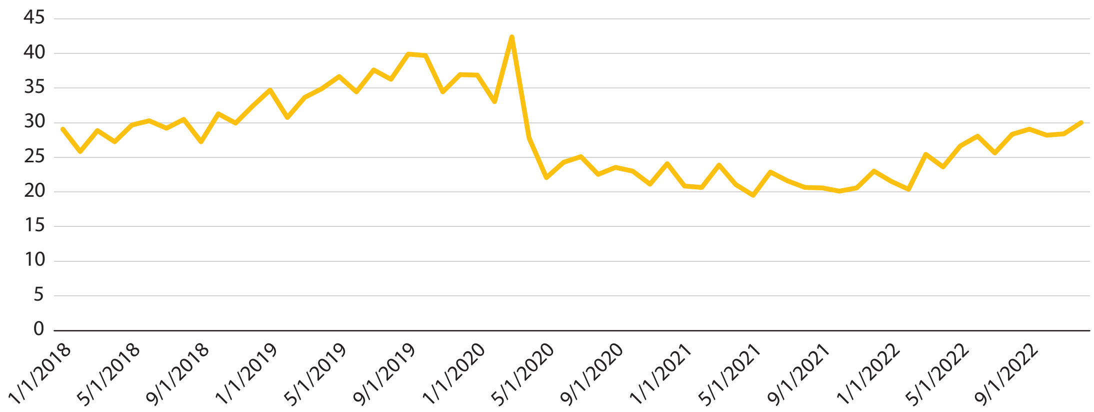
As already noted, there is an additional important caveat that must be borne in mind. If the fiscal dominance shock is a surprise to the market, the good news for government spenders is that a sudden and persistent inflation shock would reduce the nominal value of outstanding government debt, causing the government debt-to-GDP ratio to fall substantially, which implies a one-time increase in the capacity to issue new interest-bearing debt, taking pressure off of the need for inflationary money creation going forward. If this had been factored into the above examples as part of the way of funding the spending, it would have implied less of a need for inflation taxation and therefore, a lower rate of inflation. But it is also possible that fiscal dominance fears could grip the market too soon rather than too late. An early rise in the inflation risk premium would raise real rates prior to hitting the fiscal dominance constraint. A pre-fiscal dominance inflation risk premium would reduce the real value of existing debt, but also raise deficits through the higher interest on new debt and perhaps raise the debt-to-GDP ratio by reducing GDP growth (see Calomiris, 1993). As noted before, that could make hitting the constraint a self-fulfilling prophecy and also raise the amount of inflation tax needed to fund deficits. So expectational errors can cut both ways with respect to the above calculation.

It is very challenging to figure out how the *real* demands for currency and reserves would actually change if inflation rose. There is a lively debate going on right now in the US about which, if any, of the money demand concepts has a well-defined demand. Positing that there is a real demand for money (declining in our example as interest ceases to be paid on reserves) is not the same as asking whether "velocity" of reserve transactions is stable. MV = PT, or M = PT/V is always true definitionally.

Furthermore, the fact that V may move around a lot does not mean that the demand for money is nonexistent. "Velocity" is in fact highly variable as a result of PT variation. Consider the relationship between total Fedwire transfers and total reserves. First, note that the rise in Fedwire transfers reflects a big increase in the proportion of securities market transactions relative to goods-and-services-related transactions, as Peter Stella (2021) has emphasized. Driven by that boom in securities transactions, the annual value of payments effected through Fedwire, the primary US payments system, increased by

Figure 1

Monthly Fedwire Transfers / Reserves



SOURCE: Board of Governors of the Federal Reserve System, Reserves of Depository Institutions: Total [TOTRESNS], retrieved from FRED®, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/TOTRESNS. Fedwire® Funds Service; https://www.frbservices.org/resources/financial-services/wires/volume-value-stats/monthly-stats.html.

about 50,000 percent between 1957 and 2007: from $1.3 trillion to $670.7 trillion. At the same time, the reserves held as deposits by banks at the Federal Reserve, the funds used to make those payments, *fell* by about 37 percent: From $22.1 billion in 1957 to $14 billion in 2007. It's not just that velocity rose over time, it's that velocity is also highly unstable from day to day, from month to month, and year to year.

How does one model the real demand for reserves under today's policy environment? Banks need reserves today even though they are not required to hold a certain amount of reserves. The reason is that reserves are uniquely useful to clear all Fedwire transactions and remain scarce on the margin. This scarcity is clear on days when there are not enough reserves to clear all of Fedwire transactions, which has led to major headaches in global markets. Banks currently tell us that they face big penalties if their reserves fall short and are required to go to the Fed for additional reserves. In the current regime, reserves earn interest that is typically above the interest paid on Treasuries.

But the reserve demand function would change if the policy environment shifted. If interest on reserves were eliminated, the demand for reserves would likely become more stable, implying less volatility of reserve velocity. Paying market (or above-market) rates of interest on reserves makes reserves close substitutes for Treasuries: They differ only in their usefulness to execute payments, such as Fedwire transactions. In my opinion, this is a major contributor to the observed instability of the demand for reserves. Banks would seek to economize on reserves as much as possible and hold them only for the purpose of exercising transactions. The history of reserve demand in the 1930s (when Treasuries had near-zero interest rates) supports my view that the demand for reserves becomes much less empirically definable when there is little interest rate difference between Treasuries and reserves (Calomiris, Mason, and Wheelock, 2023).

Given the powerful inflation reduction that a large reserve requirement would imply (in our example, a decline from 16 percent to 8 percent), I think it is likely that a fiscal dominance shock would be accompanied by both the elimination of interest on reserves and a substantial reserve requirement (perhaps lower than 40 percent, but not so much lower that the implied rate of inflation would be permitted to rise too much).

If fiscal dominance results in a large increase in the reserve requirement, that means that the real shock to the profitability of the banking system would be greater than the shock related to the elimination of interest on reserves alone.

CONCLUSIONS

As the result of the high current US government debt-to-GDP ratio and continuing projected deficits, we face a possible dollar inflation uncertainty nightmare: Continuing deficits, if unchecked, eventually will lead to a fiscal dominance problem. This problem seems likely, given the way Congress has behaved in recent years. A significant rise in long-run real interest rates also seems quite possible, given that the three decades of decline in real interest rates are poorly understood and may reflect temporary demographic influences. Such an environment would hasten the triggering of a fiscal dominance problem, leading to a messy monetization in the US, with ramifications worldwide.

Many things would likely change in a fiscal dominance scenario to make the inflation tax base larger to facilitate the funding of continuing deficits with less of a rise in inflation. Interest on reserves would likely be eliminated—otherwise, monetization would do little to relax the constraint on the government. Inflation would rise, potentially by a large amount, if that is the only policy used to create inflation taxation. If the elimination of interest on reserves were accompanied by a new large reserve requirement, inflationary consequences could be much lower.

If the bond market does not anticipate a fiscal dominance shock sufficiently far in advance (where the definition of "sufficiently far" is determined by the duration of bonds held by the public), then bond investors would be caught with losses on high-duration bonds. All of these changes imply that the effects on banks and mutual funds and pension funds and others would be potentially quite dramatic.

In the 1970s and 1980s, major financial disintermediation from banks accompanied the rise in inflation taxation because rising inflation reduced the real rate earned on bank deposits. Similar pressures to disintermediate banks could rise again as the result of a rise in inflation taxation. If that occurs, however, banks and their political allies will redouble their efforts to use regulation to protect the banking system from innovation and competition, as they have already been doing (see Calomiris, 2021). Ultimately, the US may face a political choice between reforming entitlement programs and tolerating high inflation and financial backwardness.

What bearing does the most recent debt ceiling agreement have on the prospects for fiscal reform to avert monetization and inflation? The agreement was largely beside the point because it focused on government expenditures that are not related to Medicare, Social Security, or defense spending. Indeed, by doing so, it reinforced the view that there is no appetite for addressing the exploding deficits that are being driven by those categories of spending.

REFERENCES

Acharya, Viral 2020. *Quest for Restoring Financial Stability in India*, Sage Publications, 2020.

Beckworth, David 2023. "The Fed Has Overseen a Remarkable Transfer of Wealth from Bondholders to Taxpayers," *Barron's*, March 22, 2023.

Cagan, Philip 1956. "The Monetary Dynamics of Hyperinflation," in: Friedman, M. (Ed.), *Studies in the Quantity Theory of Money*, University of Chicago Press, Chicago, pp. 25–117

Calomiris, Charles W. 1993. "Greenback Resumption and Silver Risk: The Economics and Politics of Monetary Regime Change in the United States, 18621900," in Michael D. Bordo and Forrest Capie, eds., *Monetary Regime Transformations*, Cambridge University Press, 1993, 86132.

Calomiris, Charles W. 2021. "Chartering the FinTech Future," *Cato Journal*, Spring/Summer 2021, pp. 383-412.

Calomiris, Charles W., and Ian Domowitz 1989. "Asset Substitutes, Money Demand and the Inflation Process in Brazil," *Journal of Money, Credit, and Banking*, February 1989, pp. 7889.

Calomiris, Charles W., and Stephen H. Haber 2014. *Fragile By Design: The Political Origins of Banking Crises and Scarce Credit*, Princeton University Press.

Calomiris, Charles W., Joseph Mason and David Wheelock 2023. "Did Doubling Reserve Requirements Cause the Recession of 1937-1938? New Evidence on the Impact of Reserve Requirements on Bank Reserve Demand and Lending," Working Paper, February 2023.

Cantor, Richard M., and Frank Packer 1996. "Determinants and Impact of Sovereign Credit Ratings," *Economic Policy Review*, Vol. 2, No. 2, Federal Reserve Bank of New York, October.

Eisner, Robert 1986. *How Real Is the Federal Deficit?* New York: Simon & Schuster.

Fry, Maxwell J. 1988. *Money, Interest, and Banking in Economic Development*, Johns Hopkins.

Gurley, John G, and Edward S. Shaw 1960. *Money in a Theory of Finance*, Brookings.

Kimbrough, Kent 2006. "Revenue Maximizing Inflation," *Journal of Monetary Economics* 53, pp. 1967-1978.

McCallum, Bennett T. 1984. "Are Bond-Financed Deficits Inflationary? A Ricardian Analysis," *Journal of Political Economy* 92, pp. 123-135.

McKinnon, Ronald I. 1973. *Money and Capital in Economic Development*, Brookings Institution.

Sargent, Thomas J. 1982. "The Ends of Four Big Inflations," in Robert E. Hall, editor, *Inflation: Causes and Effects*, University of Chicago Press, pp. 41-98.

Stella, Peter 2021. "Interpreting Modern Monetary Reality," *Journal of Applied Corporate Finance* 33 (Fall): 8-23.

FEDERAL RESERVE BANK of ST. LOUIS
CENTRAL TO AMERICA'S ECONOMY®

The COVID-19 Pandemic and Inflation: Lessons from Major US Wars

Kevin L. Kliesen and David C. Wheelock

Abstract

US fiscal and monetary policies implemented during the COVID-19 pandemic have been likened to those often adopted during wars. This article compares macroeconomic policies of the pandemic period with those of major US wars since the Civil War. Inflation often surges during wars, as it did in the second year of the pandemic, and the wartime experiences can provide insights about the relative scale and persistence of inflation associated with sudden, large increases in government expenditures, such as the fiscal response to the COVID-19 pandemic. The article describes fiscal and monetary policies in each war and postwar period and traces differences in the relationships between the growth in government debt, the money stock, and inflation across the episodes to differences in the prevailing monetary regime and other institutional arrangements. The evidence from US wars suggests that the extent of government spending and the means used to finance that spending can have a significant impact on inflation outcomes. Substantial monetary financing of large increases in government spending was a characteristic of most major wars and a key driver of inflation. Further, the historical record reveals that postwar periods can be disruptive, with sharp fluctuations in economic activity and inflation, and that quick restoration of price stability requires recalibration of fiscal and monetary policy that often has been politically and technically challenging.

JEL codes: E31, E42, E51, E52, E58, E62, E64, N11, N12

Federal Reserve Bank of St. Louis *Review*, Fourth Quarter 2023, *105*(4), pp. 234-60.
https://doi.org/10.20955/r.105.234-60

> "A simple way to think about what's happened is the pandemic was like a war, and you had war financing—lots of spending, not just in the U.S. but around the world. If we were going to err, we were going to err on the side of doing too much."
> —James Bullard[1]

1. Federal Reserve Bank of St. Louis president James Bullard, quoted in Timiraos (2022).

Kevin L. Kliesen is a business economist and research officer, and David C. Wheelock is a senior vice president and special policy advisor at the Federal Reserve Bank of St. Louis. Jason Dunn and Cassandra Marks provided research assistance on this article. The authors thank Subhayu Bandyopadhyay, Bill Dupor, and B. Ravikumar for comments and George Hall for providing historical data on the Federal Reserve's holdings of Treasury securities.

The substantial increase in inflation that the United States experienced in 2021-22 was a sharp break from the previous decade. Although this higher inflation is often attributed to supply disruptions, it followed large increases in federal government spending and debt and money stock growth that reflected the fiscal and monetary responses to the COVID-19 pandemic. These responses resembled in several respects the economic policy actions taken during major wars. Parallels between the 2020-21 episode and previous war-time periods are striking: Sharp increases in government spending and inflation have broadly characterized major wars and immediate postwar periods throughout US history. However, differences in how wartime spending was financed, existing monetary regimes, and the use of wage and price controls affected the timing and extent to which inflation followed from sudden large increases in government spending. Similarly, how fast inflation retreated when wars ended, if it did at all, reflected both institutional arrangements and economic policy choices.

This article compares and contrasts economic outcomes associated with the war on COVID-19 with major US wars since the Civil War: World Wars I and II, the Korean War, and the Vietnam War. Specifically, our intent is to glean insights about the relative scale and persistence of inflation associated with, and resulting from, wartime financing—that is, sudden, large increases in fiscal expenditures. Section 1 discusses key aspects of the pandemic and its effect on the US economy. Section 2 discusses the financing of wars historically and summarizes the findings of other recent studies of US war financing. Section 3 presents evidence on inflation and the growth of government debt and the money stock across major US wars from the Civil War to the "war" on COVID-19. We highlight similarities and differences in fiscal and monetary policies associated with each war and postwar periods. Although each episode involved a substantial increase in government spending and debt, differences in the prevailing monetary regimes and other institutional arrangements contributed to differences in the relationships between the growth in government debt, the money stock, and inflation across the wartime periods, details of which we present in an appendix. This review illustrates that the US government used a considerable variety of methods to finance war expenditures while simultaneously attempting to contain inflation. Monetary financing in support of financing wars since the founding of the Federal Reserve in 1913 has been important. Section 4 concludes.

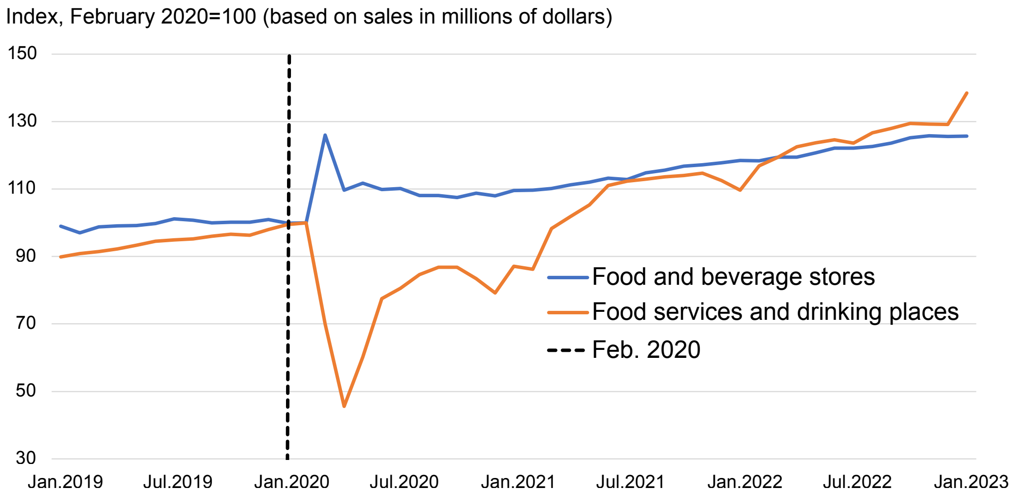
1 THE COVID-19 PANDEMIC AND THE US ECONOMY

The novel coronavirus that became known as COVID-19 emerged in Wuhan, China, in late 2019 and began to overwhelm most countries of the world in early 2020. When COVID-19 began to spread in the United States in early 2020, the federal government and the Federal Reserve responded swiftly to the emerging economic and health care crisis.[2] The Secretary of Health and Human Services declared a public health emergency on January 31, 2020. On March 13, President Trump declared a national emergency; and on March 16, the Administration issued coronavirus guidelines that encouraged the public to limit the size of gatherings and avoid restaurants and bars. Although the guidelines were intended for two weeks, the pandemic worsened dramatically, and most states enacted stay-at-home orders in March. In response, many businesses directed their employees to work remotely while other firms closed and furloughed their employees. Even in locations where businesses remained open, many individuals chose to avoid restaurants, travel, and other forms of social and economic interaction. Figure 1 illustrates this pattern: Early in the pandemic, retail sales at food and beverage stores (food consumed at home) increased sharply while sales from food services and drinking places (food consumed away from home) plunged.

2. The major actions and events are summarized on the *Timeline of Events related to the COVID-19 Pandemic* hosted on the Federal Reserve Bank of St. Louis' FRASER website (https://fraser.stlouisfed.org/timeline/covid-19-pandemic#7).

Figure 1

Retail Sales at Food and Beverage Stores and Food Services and Drinking Places, 2019-Present



SOURCE: Bureau of Economic Analysis and Haver Analytics.

Besides being a public health emergency, the pandemic was a significant economic shock, triggering a record decline in real GDP in the second quarter of 2020—29.9 percent at an annual rate. Nonfarm payrolls plunged by a seasonally adjusted 1.4 million jobs in March 2020 and then fell by an astounding 20.5 million the following month. The civilian unemployment rate surged from a 50-year low of 3.5 percent in January and February 2020, to a more than 80-year high of 14.7 percent in April 2020.[3] The inflation rate (measured as 12-month percent changes) fell from 1.8 percent in January 2020 to 0.4 percent in May 2020. In the Federal Open Market Committee's June 2020 Summary of Economic Projections (SEP), the median FOMC participant projected that both headline and core PCE price inflation would remain under the Committee's 2 percent target through the end of 2022.

In relatively short order, the National Bureau of Economic Research (NBER) declared that a recession had begun in February 2020. The recession was short-lived, however, as real GDP growth rebounded at a 35.3 percent annual rate in the third quarter of 2020 and the unemployment rate dropped by more than half to 6.7 percent by year end. By early 2021, aggregate real personal income exceeded what it would have been if income had simply grown at its pre-pandemic trend rate (Bullard 2021). The NBER subsequently determined that the recession had ended in April 2020. The pandemic-spawned recession was the shortest, but deepest, US recession on record.[4]

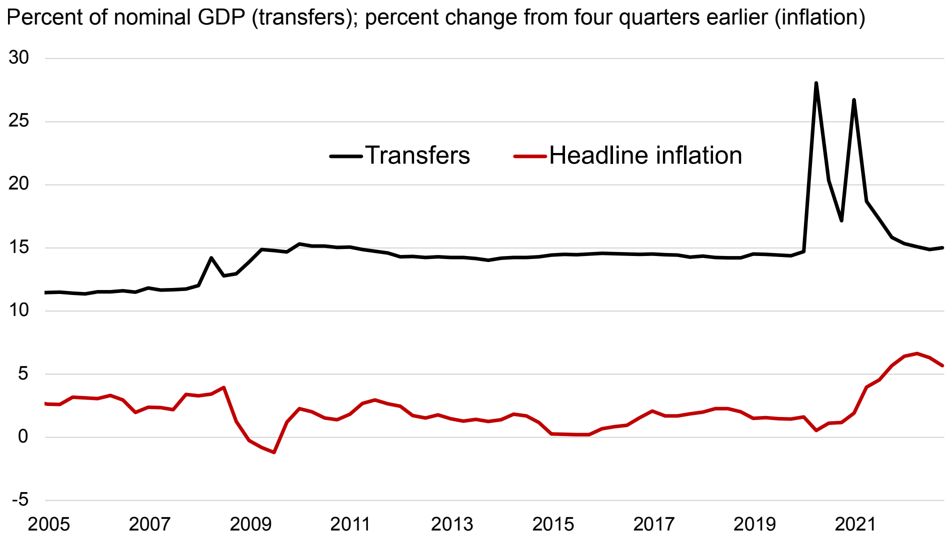
In response to the evolving economic and health care crisis, four pandemic-specific pieces of legislation were signed into law during the spring of 2020. The total amount allocated by Congress at the time exceeded $2.7 trillion, with the CARES Act alone accounting for $2.2 trillion of the total spending. The CARES Act included approximately $450 billion to fund five special Federal Reserve (Fed) lending facilities that were established to support the financial system. Subsequently, in early 2021, Congress enacted two additional pieces of legislation, including the American Rescue Plan Act (ARPA), a $1.9 trillion spending package intended to further encourage recovery from the economic and public health

3. According to the Bureau of Labor Statistics, the actual unemployment rate would have exceeded 19 percent if respondents had correctly reported their pandemic-related employment status. See https://www.bls.gov/news.release/archives/empsit_05082020.pdf.

4. The annual report of the Federal Reserve Bank of St. Louis for 2020 describes some economic effects of the pandemic on the U.S. economy (https://www.stlouisfed.org/annual-report/2020).

Figure 2

Federal Government Transfer Payments and PCEPI Inflation, 2005-22



NOTE: Data through 2022:Q4.

SOURCE: BEA and Haver Analytics.

impacts of the pandemic. In all, the six major pieces of federal legislation enacted in 2020 and 2021 resulted in a more than $5 trillion increase in the nominal federal budget deficit over 10 years.[5]

The Fed's monetary policy response to the pandemic was similarly aggressive. On February 28, 2020, Federal Reserve Chair Jerome Powell issued a statement declaring that "the coronavirus poses evolving risks to economic activity."[6] Two weeks later, on March 15, the FOMC held an unscheduled meeting and voted to lower the federal funds target rate by 100 basis points to 0 to 0.25 percent. In addition to lowering its federal funds rate target effectively to zero, the Fed's Open Market Committee (FOMC) initiated purchases of Treasury and mortgage-backed securities totaling $120 billion per month. From the fourth quarter of 2019 to the fourth quarter of 2021, the value of assets on the Fed's balance sheet increased from $4.17 trillion (19.2 percent of GDP) to $8.76 trillion (36 percent of GDP).

At a high level, the fiscal and monetary response to the COVID-19 pandemic resembled in several respects the economic policy actions taken during previous major wars reviewed in this article. However, unlike most wars, when governments spend large amounts on armaments and mobilizing troops, the appearance of an enemy in late 2019 (the virus) triggered a massive increase in federal spending in 2020-21 that consisted largely of transfer payments to households and firms most directly impacted by business shutdowns and other measures taken to slow the spread of the virus. Figure 2 shows that federal transfer payments averaged 14.5 percent of GDP from the first quarter of 2019 to the first quarter of 2020. From the second quarter of 2020 through the second quarter of 2021, when transfer payments peaked at 28.1 percent of GDP, they averaged 22.2 percent.

5. The four pieces of legislation enacted in 2020 were (1) the *Coronavirus Preparedness and Response Supplemental Appropriations Act*; (2) the Families First Coronavirus Response Act; (3) the *Coronavirus Aid, Relief, and Economic Security (or CARES) Act*; and (4) the Paycheck Protection Program and Health Care Enhancement Act. In 2022, the following two pieces of legislation were enacted: (1) the *Consolidated Appropriations Act* and (2) the *American Rescue Plan, or ARPA*. These are listed and discussed in Congressional Research Service (2021). Per convention, the budgetary impacts are reported over a 10-year period (see Congressional Research Service, 2021, Table 2) for budgetary impact details.

6. Statement from Federal Reserve Chair Jerome H. Powell, February 28, 2020 (Federal Reserve Board – Statement from Federal Reserve Chair Jerome H. Powell).

The increased spending was financed primarily by borrowing. The federal budget deficit rose from $984 billion (4.6 percent of GDP) in fiscal year 2019 to $3.1 trillion (15 percent of GDP) in 2020. Helped by a sharp acceleration in real GDP growth in 2021, the budget deficit declined modestly in fiscal year 2021 to $2.8 trillion (12.3 percent of GDP).

The massive monetary and fiscal response triggered an acceleration in the aggregate demand for goods and services that, combined with temporary supply-side disruptions that lowered potential real GDP, provided a powerful inflation impulse.

Initially, few forecasters and economists drew parallels between the fiscal and monetary actions in response to the economic fallout of COVID-19 and those of past wars. This may have been because inflation, as measured by the personal consumption expenditures price index (PCEPI), remained tame through early 2021. The headline (all items) PCEPI increased by just 1.3 percent in 2020 (December to December), well below the Fed's 2 percent target. Measured on a 12-month basis, PCEPI inflation (both headline and core) remained below 2 percent in January and February 2021. Nonetheless, FOMC members expected a small, temporary increase in the rate of inflation in 2021. In the FOMC's March 2021 Summary of Economic Projections, the median participant projected that PCEPI inflation would be 2.4 percent in 2021, 2 percent in 2022, and 2.1 percent in 2023.

Contrary to most predictions, inflation began to rise sharply beginning in March 2021 and continued to increase until the middle of 2022. Some economists, such as Summers (2022), began to warn of the inflationary impact of the COVID fiscal programs. However, the magnitude and duration of the inflation surge of 2021 and 2022 was a surprise to many policymakers and economists (Waller, 2022). Higher inflation was widely expected to be short-lived, reflecting supply-side disturbances, higher energy shocks, and a temporary increase in aggregate demand spurred importantly by record-high saving balances that had accumulated during the pandemic.[7] In short, most policymakers and forecasters expected inflation to return to its pre-pandemic pattern once the shocks dissipated. For example, the Council of Economic Advisers (2021) likened the increase in inflation to the temporary burst of inflation that followed the removal of price controls at the end of World War II.[8] Many Fed officials agreed.

With inflation beginning to trend higher in the spring of 2021, the median FOMC participant in the June 2021 Summary of Economic Projections projected that PCEPI inflation would only rise to 3.4 percent in 2021 and then slow to 2.1 percent in 2022. Speaking at Jackson Hole, Wyoming, in August 2021, Chair Powell claimed that high inflation was unlikely to persist and that the FOMC's near-term focus would be on achieving maximum employment.[9] Inflation continued to increase through the remainder of 2021, however, and by December 2021 the 12-month percent change in the PCEPI inflation rate had reached a 40-year high of 6.0 percent; the CPI inflation rate was even higher at 7 percent, and thus inflation was triple the Fed's 2 percent inflation target.

When the FOMC announced a decision to taper asset purchases on May 4, 2022, CPI inflation had surpassed 8 percent and the PCE price inflation rate exceeded 6 percent (both measured on a 12-month basis). Inflation peaked in June 2022 at 9.1 percent measured by the CPI and at 7 percent measured by the PCEPI. By the end of 2022, measures of headline inflation had retreated—largely reflecting lower energy prices. At that time, the key question for FOMC policymakers was how much additional policy

7. Recent research from New York and San Francisco Fed economists offers divergent evidence on the relative contributions of supply and demand to inflation. Giovanni (2022) finds that 60 percent of the inflation over the 2019-2021 period stemmed from increased demand for goods and services, with the remaining 40 percent due to supply-side issues. Using a shorter time span, Shapiro (2022), by contrast, finds that supply factors explained about half of the increase in inflation over the 12 months ending in April 2022, with demand factors explaining about a third of the increase. Using a fiscal theory of the price level, model-based approach, Bianchi and Melosi (2022) argue that ARPA accounted for 3.5 percentage points of the recent increase in inflation, with supply-side (cost-push) factors accounting for a similar percentage increase.

8. The Council of Economic Advisors report, "Historical Parallels to Today's Inflationary Episode," is dated July 6, 2021 (https://www.whitehouse.gov/cea/written-materials/2021/07/06/historical-parallels-to-todays-inflationary-episode/).

9. See https://www.federalreserve.gov/newsevents/speech/powell20210827a.htm.

tightening would be required to return inflation to the Committee's 2 percent target. High inflation remained a threat to macroeconomic stability and to the FOMC's credibility in 2023. Accordingly, as indicated by the minutes of the January 31–February 1, 2023, FOMC meeting, monetary policymakers remained firmly focused on returning inflation to the 2 percent target.

The linkages between sudden large increases in government spending and high inflation are not new and were a feature of major US wars of the past. The next section discusses some historical comparisons. The post-COVID inflation dynamics studied in this article suggest that an end to wartime financing eventually results in lower rates of inflation, but the transition has not always been smooth.

2 FINANCING WARS: AN HISTORICAL PERSPECTIVE

Historically, wars have been characterized by heavy government spending and debt issuance, highly accommodative monetary policy, and often inflation (e.g., Rockoff 2015). Tightening monetary policy to control inflation during wartime could be viewed as inconsistent with the aim of winning the war, especially if doing so would entail higher interest rates and thus higher government borrowing costs. Consequently, governments have often accepted inflation (or imposed controls to limit price increases) with or without the acquiescence of their central banks to ensure sufficient low-cost funding of war expenditures. Situations in which monetary policy becomes subservient to fiscal policy, referred to as "fiscal dominance," seem more likely during wars or other emergencies that cause governments to borrow heavily to finance large increases in spending. "Fiscal dominance" may be defined as "the imposition of the fiscal authority's preferences despite disagreement from the central bank." In a fiscal dominance regime, "the preferences of the central bank, and thus its independence, [is] irrelevant" (Martin 2021, p. 3). In this article, we use a less strict definition of fiscal dominance as shorthand to refer to situations in which monetary policy is determined primarily by the fiscal authority regardless of agreement by the monetary authority.

Whether fiscal dominance persists after the emergency passes could depend on a variety of conditions, including the level of outstanding government debt and degree of central bank independence.[10] Our review of wartime episodes indicates that higher inflation has been a characteristic of most major US wars and was associated with monetary accommodation of sudden, large increases in government expenditures. Further, we find that the restoration of price stability after wars has always involved a tightening of both fiscal and monetary policy but that features unique to each episode influenced the timing and extent of economic disruption associated with each disinflation. For interested readers, the appendix provides more detailed discussions of fiscal and monetary actions, and the behavior of inflation, during each major war reviewed in this article.

Governments can finance expenditures for wars and other projects by raising revenues (e.g., taxes) or by issuing debt. Debt sold to the public can be either interest-bearing or non-interest-bearing and may have qualities that enable it to function as money. Economists have long argued about the optimal financing of wars and other emergency expenditures. Adam Smith, for example, asserted that borrowing by issuing long-term bonds hides the true funding cost of wars and thereby provides an incentive for political leaders to wage war. By contrast, John Stuart Mill contended that borrowing is acceptable so long as it does not cause interest rates to rise. The modern view of optimal war finance stems from Barro (1979), who argues that governments should borrow to finance temporary increases in government expenditures to minimize disincentive effects associated with higher taxes.[11]

10. According to Schnabel (2020), "At the time of the Maastricht Treaty, high government debt was seen as a major threat to central bank independence, and it was feared that *fiscal dominance* could induce a central bank to deviate from its monetary policy objectives, endangering price stability" (emphasis in original).

11. However, Lucas and Stokey (1983) argue that in an economy with state-contingent policies, "the efficient policy is to tax capital heavily or adjust returns on government debt immediately upon the outbreak of war" (Ohanian 1997, p. 23). Hall and Sargent (2021) elaborate on the differences between the Barro (1979) and Lucas and Stokey (1983) frameworks and conclusions.

Before the development of modern capital markets, governments sought to finance military expenditures by levying taxes (broadly defined to include impressment of property and soldiers). As noted by Keen and Slemrod (2021), examples from several centuries of world history include the "Saladin tithe" to finance a 12th century crusade and the "fifteenth and tenth" tax on the value of movable goods, such as corn or farm animals, in fourteenth century England. Later, in the 16th century, Henry VIII levied a class-based tax whereby dukes paid a larger tax than barons, who in turn paid more than peasants. However, governments resorted to borrowing when fully financing expenditures through taxes proved impossible or undesirable, and they used various mechanisms to increase the public's willingness to accept the government's debt. Some of those mechanisms included issuing debt that either directly or indirectly augmented the supply of money, for example by declaring the debt securities to be legal tender. Other funding mechanisms involved borrowing from banks, sometimes in exchange for special privileges that gave rise to the first central banks. The Bank of England, for example, was founded in 1694 during a war with France. Kindleberger (1984, p. 5) writes that "It is no accident, for example, that the Bank of England was established in the midst of the Nine Years' War, … or that the Bank of France was established by Napoleon in 1800 to help finance his wars."[12] The founders of the Bank of England agreed to lend £1.2 million to the English government in exchange for a £100,000 annual payment and a monopoly right to issue bank notes in London. The Bank's note issues, according to Kindleberger (1984, p. 53), had the effect of exacerbating wartime inflation.[13]

In his review of the financial histories of US military conflicts, Rockoff (2015) finds that money creation was a significant component of financing "major" US wars but not of "minor" wars, such as the Mexican-American War of the 1840s and Gulf War of 1991. Not surprisingly, therefore, inflation was a characteristic of major wars but not of minor conflicts. However, funding mechanisms and inflation outcomes differed somewhat across even the major wars, depending on the prevailing monetary regime, as well as on the timing and extent of price controls and other institutional arrangements.

The extent to which governments borrow to finance wartime expenditures can depend at least in part on their ability to raise revenues by other means. Borrowing by issuing debt that serves as a means of payment, either directly or via the banking system, is often termed "printing money" or use of the "inflation tax" because of the association between increases in the supply of money and inflation. An extreme example of "printing money" to finance a war was by the Confederate States of America during the American Civil War. The Confederacy financed approximately 60 percent of its spending by printing money and incurred an inflation rate that exceeded 100 percent per year from mid-1862 through the end of the war in 1865.[14]

Whereas some governments have little choice but to attempt to finance war expenditures by printing money, the funding method chosen can also reflect the preferences of those in charge. The United States financed its Korean War expenditures mainly by raising taxes, for example, reflecting President Harry Truman's aversion to deficit spending and high interest rates (Ohanian 1997). Hall and Sargent (2021) study the financing of World Wars I and II and the "War on COVID-19." They distinguish between the issuance of interest-bearing debt (i.e., Treasury securities) and non-interest-bearing debt (i.e., base money).[15] Hall and Sargent find that borrowing was the preferred method in all three wars; taxes were raised to some extent in WWI and WWII, but hardly at all in the COVID-19 episode. A key feature of the COVID-19 period was that the monetary base rose sharply alongside the increase in

12. Based on nearly a century of data from the London bond market, Poast (2015) finds that having a central bank lowered a sovereign's borrowing costs—particularly during wartime.

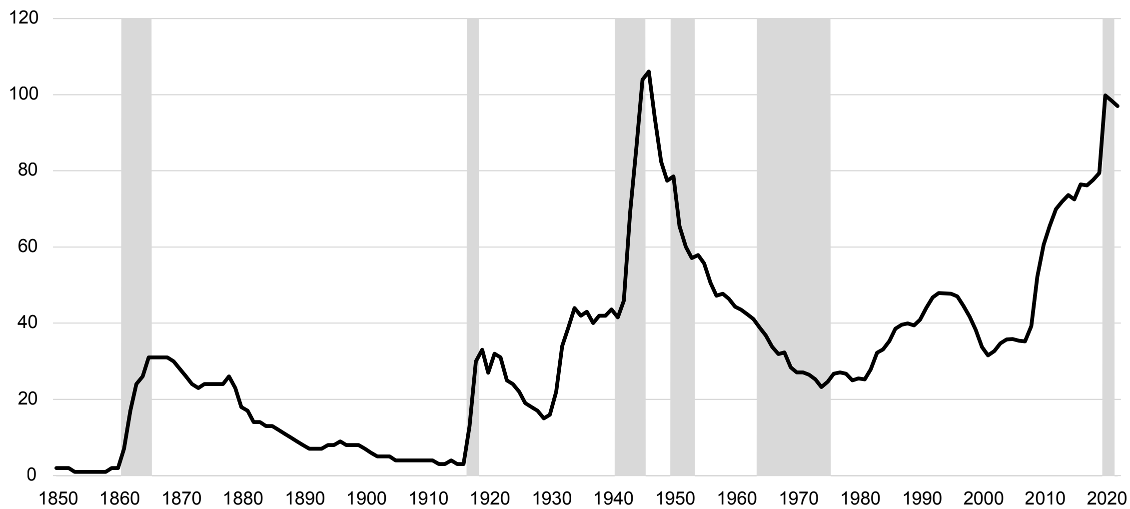
13. See also https://www.bankofengland.co.uk/about/history.

14. For additional information about Confederate war finance and inflation, see Weidenmier (https://eh.net/encyclopedia/money-and-finance-in-the-confederate-states-of-america/) or *Tax History Project -- The Civil War* (www.taxhistory.org).

15. The monetary base reflects the monetary liabilities of the central bank, i.e., currency outstanding and the reserve deposits of banks and other depository institutions.

Figure 3

US Federal Debt Held by the Public as a Percent of GDP During Major US Wars, 1850-2022



NOTE: Shaded areas are the war years studied in this paper.

SOURCE: Congressional Budget Office.

government expenditures as the Federal Reserve purchased Treasury securities in the open market equal to about half of the increase in gross federal debt.[16]

3 MAJOR US WARS, GOVERNMENT DEBT, AND INFLATION

Inflation has occurred during nearly every major US war, typically the result of monetary accommodation of wartime government spending. But history also shows that monetary accommodation can take different forms. These have included simply issuing new currency to pay for government purchases, using financial incentives to encourage the public to purchase government debt, and outright purchases of substantial amounts of government debt by the central bank. The different mechanisms can make it difficult to discern the extent of central bank financing from basic data on the size of the Fed's balance sheet, for example. And, similarly, price controls can obscure relationships between money growth and inflation. This variation can be explained to a great extent by differences in institutional features affecting the relationships between government debt, money, and inflation during the wars and early postwar periods, which we discuss in this section and the appendix.

Figures 3 and 4 provide historical data on US federal debt held by the public from 1850 to 2022 (as a percent of nominal GDP) and inflation. Inflation is measured as the annual percentage change in the consumer price index (CPI).[17] The shaded areas are the years of the major wars discussed in this article and the COVID-19 period, which we date as the years 2020 and 2021. Although COVID-19 deaths and hospitalizations persist today, the main elements of the fiscal and monetary response to the pandemic occurred in 2020 and 2021.[18]

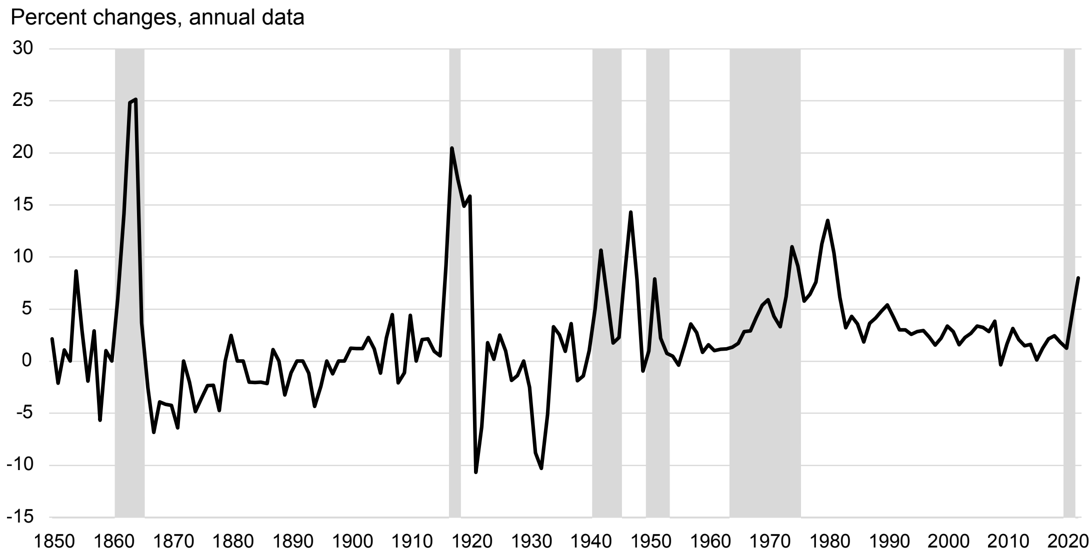
16. Gross federal debt includes debt held by the public and debt held by government trust funds (e.g., Social Security and Medicare) and other government accounts. Debt held by the public includes debt on the Federal Reserve's balance sheet. Federal debt held by the public increased by $4.2 trillion from 2019 to 2021, and the Federal Reserve's holdings of Treasury securities increased by $3.3 trillion. Hence, the Fed purchased 78.8 percent of the increase in federal debt held by the public (data are calendar year averages).

17. See the Data Appendix for source information.

18. We do not claim that any structural effects on the economy from the COVID-19 pandemic ended after 2021. There is a growing literature on the economic effects of COVID-19, particularly on the supply and demand for labor and global supply chains.

Figure 4

CPI Inflation During Six Major US Wars, 1850-2022



NOTE: Inflation rates are calculated from annual average data. Shaded areas are the war years studied in this paper.
SOURCE: www.measuringworth.com and the Department of Veterans' Affairs.

Several patterns are evident in Figures 3 and 4. First, relative to GDP, federal debt held by the public increased sharply in the first three wars, but then fell when the wars ended. This pattern did not hold for the Korean and Vietnam wars, however, as debt was falling relative to GDP before those wars and continued to fall during them. Second, as Figure 3 shows, there were two non-war periods when federal debt rose sharply relative to GDP. The first occurred during the 1930s when GDP dropped sharply during the first years of the Great Depression, causing the debt-to-GDP ratio to rise despite little change in debt outstanding. The second occurred during the financial crisis in 2008 and subsequent recession. Figure 3 shows that the debt-to-GDP ratio rose sharply after the financial crisis and that, when the pandemic hit, debt as a percent of GDP was already at its highest level since just after World War II. Debt as a percent of GDP then rose sharply during the pandemic. Thus, the COVID-19 era most closely resembles the first three wars (Civil War, World War I, and World War II).

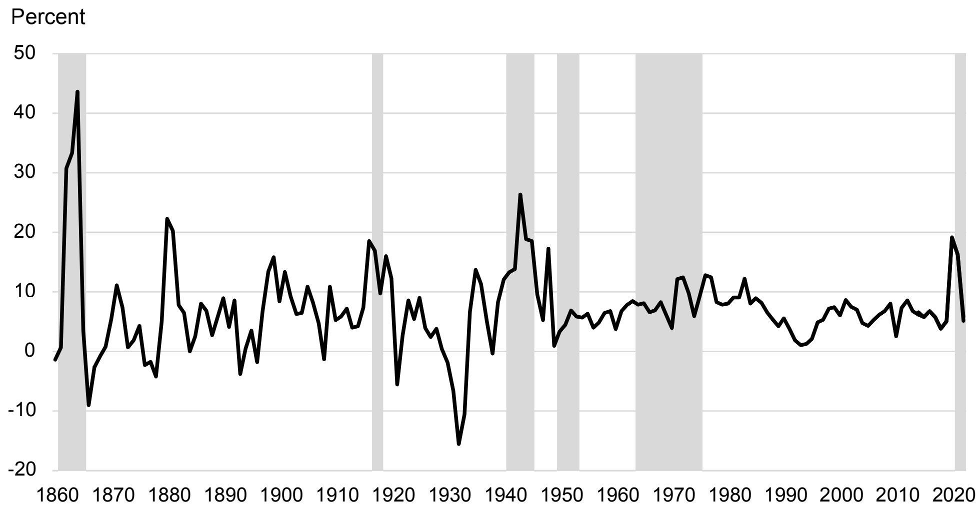
The CBO estimates that the six legislative actions signed into law in 2020-21 increased the federal budget deficit by $5.04 trillion over a 10-year period (Congressional Research Service, 2021). As shown in Figure 3, federal debt held by the public as a percent of nominal GDP increased from 79.4 percent in 2019 to 99.8 percent in 2020. Thereafter, the debt-to-GDP ratio fell to 98.4 percent in 2021 and then slightly more to 97 percent in 2022.[19]

There is a long literature on the connection between expansionary fiscal policy and inflation (e.g., Bordo and Levy, 2020). Figure 4 plots CPI inflation since the Civil War. Inflation rose sharply during the major wars, but then decelerated quickly after the wars ended (with the exception of the Vietnam War). The ends of the Civil War and World War I were also followed by periods of significant deflation, but that was not true of later wars. Finally, as with the previous five wars, CPI inflation spiked during the COVID-19 pandemic and remained at its highest levels since the early 1980s in 2021 (6.7 percent) and 2022 (7.1 percent). According to the median FOMC participant in the March 2023 Summary of Economic Projections (SEP), inflation is not expected to return to the Committee's 2 percent inflation target until sometime after 2025.

19. See https://fred.stlouisfed.org/series/FYGFGDQ188S.

Figure 5

Growth of the M2 Money Stock and Major Wars, 1860-2022

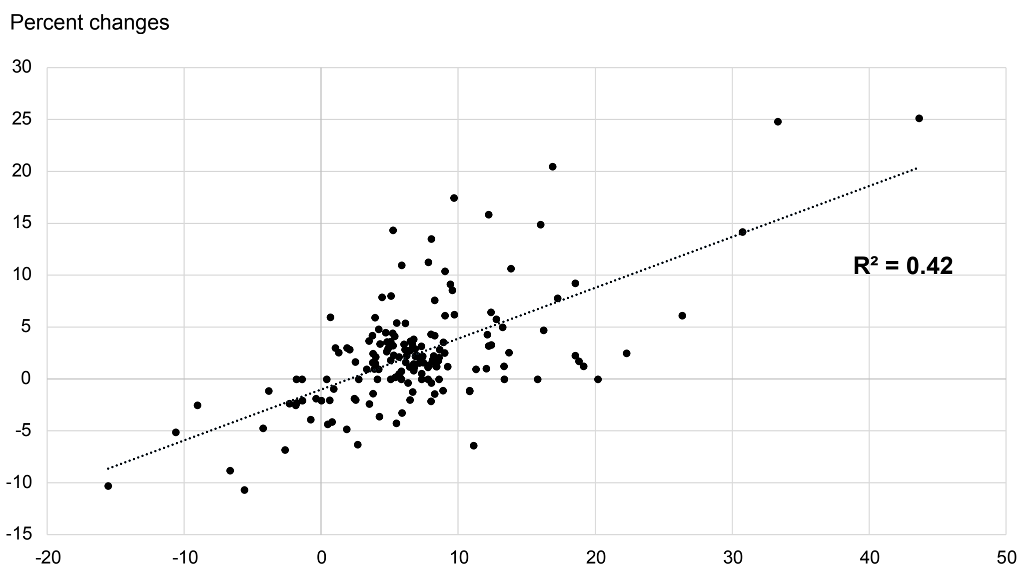


NOTE: Shaded areas are the war years studied in this paper.

SOURCE: Various (see text for data sources).

Figure 6

Money Growth and CPI Inflation, 1860-2022



SOURCE: Data from Figures 2 and 5.

History has shown that sustained, high rates of inflation cannot typically occur without excessive money growth. Figure 5 plots the growth rate of the money stock for the wartime periods studied in this article. Except for the Korean War, the major US wars were associated with marked increases in money stock growth for the duration of the war. The COVID-19 era has not been exceptional in that regard. As Figure 5 shows, however, the extent to which the money stock increased varied considerably across the episodes. Further, as with the pattern for inflation, the money stock contracted sharply after the Civil War and World War I, but not after the more recent wars. The differences between the first

Table 1

Major US Wars Since the Civil War: Key Facts

War	Period of hostilities	Debt/GDP avg prior four years	Debt/GDP avg during war	Years of fiscal dominance	Average inflation rate	Years of price controls***	Fed share of marketable Treasury debt (beginning)	Fed share of marketable Treasury debt (end)	Fed share of marketable Treasury debt (peak)
Civil War	1861-65	1.5	21.0	1861-65	14.8	N/A	N/A	N/A	N/A
WWI	1917-18	3.3	21.5	1917-19	19.0	1917-18	4.9	0.67	9.51
WWII	1941-45	41.9	60.8	1941-51*	5.9	1942-46	5.49	9.42	9.51
Korean	1950-53	90.0	65.3	1950-51*	3.0	1951-53	11.22	16.24	16.93
Vietnam	1964-75	42.8	29.6	N/A	4.9	1971-74	16.93	26.63	29.93
COVID-19**	2020-21	77.4	100.0	N/A	3.0	N/A	14.26	24.38	24.71

NOTE: *Fiscal dominance period ended with the Fed-Treasury Accord in March 1951. **For the COVID-19 period, we use annual data for debt/GDP and CPI inflation from the CBO and BLS. The Fed share of marketable debt is annual averages from monthly data from Hall and Sargent (2022). ***Price controls were in place Aug. 1917–Nov. 1918; April 1942–June 1946; Jan. 1951–Feb. 1953; Aug. 1971–April 1974 (Rockoff, 1981).

SOURCE: Bureau of Economic Analysis, Bureau of Labor Statistics, Congressional Budget Office, MeasuringWealth.com, and Hall and Sargent (2022).

two wars and the subsequent wars likely reflect the different monetary regimes of the eras. Under the gold standard that prevailed in the 19th and early 20th centuries, the price level was volatile in the short run but mean-reverting over time (Bordo 1981). Under the monetary regimes that have prevailed since World War II, however, the price level has exhibited less volatility but has not been mean-reverting. This suggests that the long-run price level has been less predictable since World War II.

Similar to the earlier wars, money stock growth began to slow in 2022. Although M2 rose by 5 percent in 2022, the money stock peaked in March 2022 and declined over the final nine months of the year. The decline after March 2022 coincided with the FOMC's tightening of monetary policy, as reflected in a 4.25-percentage-point increase in the federal funds rate target between March and December 2022. Borio et al. (2022) present a cross-country regime-based model and show that periods of high inflation correspond to periods of excess money growth (the difference between money growth and real GDP growth). This finding correlates well with the surge in money growth during the COVID-19 period. As Figure 6 shows, even over longer periods of time, the correlation between the annual money growth rates and annual CPI inflation rates (plotted in Figures 2 and 5) is relatively strong.[20]

Table 1 provides data on the increases in government debt (relative to nominal GDP), changes in the share of Treasury debt held by the Federal Reserve, and inflation during the five major wars examined here and the COVID-19 period. During the first three wars, the debt-to-GDP ratio rose sharply relative to the average over the previous four years. That ratio rose sharply during World War II, although, as Figure 3 shows, the average of 60.8 percent belies the eventual peak of more than 100 percent. Federal debt outstanding more than quintupled, rising from $51 billion in 1940 to $260 billion in 1945 (Carter et al., 2006, series ea682). Public debt outstanding eventually peaked a year after the war ended, in 1946, at $271 billion (106.1 percent of GDP). Through 2022, the debt-to-GDP ratio never surpassed the record debt level of 1946; but the debt level of 100.3 percent in 2020 did come close (see Figure 1). Debt that was accumulated during the post-World War II years was subsequently reduced by fiscal tightening, economic growth, and inflation, as reflected in the declining debt-to-GDP ratio throughout the pre-war and wartime years of the Korean and Vietnam periods.

Table 1 shows that inflation was a characteristic of each wartime episode, as were price controls, though the experiences varied. As noted above, wartime inflation was highest during the Civil War

20. The linear regression line shown in Figure 6 is based on a least-squares regression between the two variables, plus a constant term.

and World War I eras, but much lower in later wars. As discussed in detail in the Appendix, temporary wage and price controls were implemented during the 20th century wars. Wage and price controls appeared to have been much more effective in keeping inflation low during the second World War and the Korean War than in World War I. However, inflation accelerated after the price controls were lifted in all three episodes. For example, after price controls were lifted in 1946, the CPI inflation rate approached 20 percent during the first half of 1947 and averaged 14.5 percent between July 1946 and December 1947. Inflation then fell sharply and between January 1949 and June 1950, the CPI inflation rate averaged -1.0 percent, i.e., a *deflation* rate of 1.0 percent. Wage and price controls were also implemented in August 1971 as part of the Nixon Administration's strategy to contain inflation (see Table 1). But unlike prior wartime experiences, inflation rose while the controls were in place and rose still higher after controls were lifted.[21] In the latter part of the Vietnam War (August 1971 to April 1974), the CPI annualized inflation rate was 4.6 percent in the six months before controls were implemented; inflation rose to 6.2 percent during the control period, and then to 12.1 percent six months after the controls were lifted (Rockoff, 1981, Table 1).

Eras of Fiscal Dominance

As noted above, fiscal dominance occurs when the monetary authority no longer exercises full control over its balance sheet. This occurred in World War II when the Fed acquiesced to the Treasury by pegging the value of short- and long-term interest rates to reduce the government's interest cost (see the Appendix for details). Fiscal dominance can lead to higher inflation as the central bank's liabilities expand to accommodate increased government spending (see Sargent, 1982; Cochrane, 2022b). For example, central bank purchases of government securities typically increase the reserves of the banking system and, hence, the monetary base. During World War II, the Fed purchased a large share of the increase in Treasury debt. However, in World War I, the Fed financed wartime expenditures mainly by lending on favorable terms to banks against their holdings of government securities. In that instance, reserves and the monetary base rose sharply even though the Fed did not directly purchase large amounts of government securities.

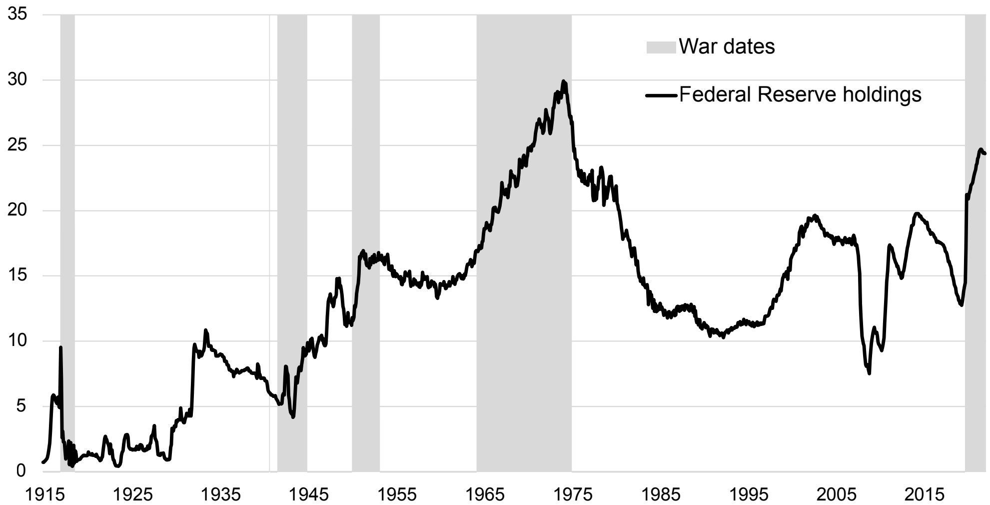
Figure 7 shows the Fed's holdings of marketable interest-bearing Treasury debt as a percent of total marketable interest-bearing Treasury debt. All else equal, an increasing share of Treasury debt held by the Fed will tend to increase the money supply and thus contribute to inflation. From 1913 to 1974, the Fed's share of marketable interest-bearing debt steadily increased, from near zero to roughly 30 percent. Thereafter, the share fell until the pandemic, but on a volatile path that exhibited no discernible trend. The figure shows, however, that the Fed's holdings of Treasury debt usually increased during wartime periods and that the Fed's holdings of interest-bearing Treasury debt rose sharply during the COVID-19 period. As noted previously, Hall and Sargent (2022) find that the Fed financed roughly half of the increase in federal debt during the war on COVID-19.

Table 1 provides a more granular analysis of debt/GDP, wartime inflation, and fiscal dominance. Fiscal dominance clearly prevailed at times in the first four wars, but not in the Vietnam War. Unlike World War I and II, the Fed did not explicitly agree to support government spending during the Vietnam War. However, Fed Chair William McChesney Martin faced considerable political pressure to adapt monetary policy to the Administration's policy objectives. He was protective of the Fed's independence and was famously taken to task by President Johnson for raising the Fed's discount rate in 1965 (Bremner 2004, pp. 208-10). However, Martin also believed in consensus and accepted limits on the Fed's independence. According to Meltzer (2009, pp. 48-49), Martin's view of Fed independence

21. George Shultz, who administered price controls in the Nixon Administration, argued that the price controls encouraged easier monetary policy, which might explain why inflation continued to rise. See Meltzer (2009, p. 761).

Figure 7

Federal Reserve Holdings of Marketable Interest-Bearing Federal Debt as a Percent of Total Marketable Interest-Bearing Debt (Par Value), 1915-2021



SOURCE: Hall and Sargent (2022).

was that of "independence within the government, not independent of the government." Further, Martin "made it clear that to him independence did not permit the Federal Reserve to prevent inflation if the administration and Congress ran large budget deficits" (Meltzer 2009, p. 85).[22] A charitable reading of the Martin tenure is that, while the Treasury never demanded that the Fed buy securities issued for wartime finance purposes, as in World Wars I and II, the seeds of the Great Inflation were sown during his tenure.

Similarly, during the COVID-19 period, the Fed did not explicitly align its monetary policies with the government's fiscal policy. Nonetheless, monetary policy in 2020-21 was aggressive and has been likened to a fiscal "helicopter drop" (see Cochrane, 2022a). However, the war on COVID-19 was not a period of fiscal dominance in the sense that the Fed opposed (overtly) the government's spending plans but was powerless to resist. Still, the pandemic was a period like other major wars since 1913 when a vigorously expansionary monetary policy coincided with a sudden, large increase in government spending. In that sense, monetary and fiscal policy acted in tandem—and maybe that is a distinction without a difference.

4 CONCLUSION

The major US wars since the Civil War were all characterized by sharp increases in government spending, rising federal debt levels, significant monetary accommodation, and inflation. We find that fiscal dominance, which can be thought of as monetary support of government spending, was a feature of most of the major wars reviewed here. The "war" on COVID-19 was also characterized by exceptional monetary accommodation, and it is perhaps surprising that more economists and policymakers were not attuned to the possibility that inflation could move sharply higher. Of course, the response to the economic impact of COVID-19 was also different from the war experiences in various ways. For example,

22. Gardner Ackley, who chaired the Council of Economic Advisors under President Johnson, reported that "on several occasions … Martin has persuaded the Open Market Committee to go along with something 'because the President wanted it.'" (Quoted by Woolley 1984, p. 116).

government spending during the pandemic consisted largely of transfer payments to households and firms to help them weather unemployment and income loss, rather than of purchases of war material and payments to service members. And, unlike most wars, which cease with an armistice, the pandemic has not had a clear end. Still, the tempering of the pandemic, the economic recovery, and rising inflation in 2021 refocused the Fed's attention on restoring price stability. It remains to be seen whether inflation will continue to decline toward a level consistent with price stability as it did after the major wars other than the Vietnam War.

Although the economic contours of the major US wars reviewed in this article had features in common, they had differences as well—for example, in the extent to which government spending was financed by "printing money," a form of fiscal dominance. Inflation outcomes differed, too, in part because of differences in the use of price controls, but also because of institutional features and changes that affected inflation expectations, especially in postwar periods, such as the Fed-Treasury Accord of March 1951. Although the inflation of the 1970s raised doubts about the ability of independent central banks to contain inflation and revived interest in restoring the gold standard, the disinflation of the 1980s and subsequent "Great Moderation" that followed provided strong evidence that a politically independent central bank can contain inflation.

The inflation of the Vietnam War era, which moved even higher after the war, is perhaps the most puzzling of all the major US war periods reviewed in this article. The Great Inflation is almost universally blamed on the Fed's monetary policy.[23] Explanations for why the Fed failed to contain inflation include the failure of policymakers to understand the fundamental cause of inflation and their unwillingness to tighten policy if that meant higher unemployment. However, the Fed was also subject to political pressure during the Johnson and Nixon Administrations. Although political pressure was less overt under the Ford and Carter Administrations, federal budget deficits increased in the second half of the 1970s and the Fed was criticized for raising interest rates while also not reducing inflation. Only when inflation became intolerable and President Carter appointed Paul Volcker as Fed chair was monetary policy used to bring inflation under control. Thus, unlike the prior war eras, the end of the Vietnam war brought neither a significant reduction in fiscal deficits nor an end to pressure on the Fed to keep interest rates from rising.

The fiscal and monetary policies adopted in the "war" on COVID-19 were indeed like past wars. Sharp increases in government spending with substantial monetary accommodation promoted a rapid recovery from the severe, but brief, recession in the second and third quarters of 2020. However, as in past wars, those fiscal and monetary policies also contributed to a significant increase in inflation. Whether price stability will be restored quickly (as it was after the Civil War and World War I or soon after the removal of price controls at the end of World War II) or persist for several years (as it did after the Vietnam War) remains to be seen. The era of modern central banking shows that central banks are capable of achieving low inflation, though the evidence from US wars suggests that the extent of government spending and the means used to finance that spending (e.g., monetizing the debt) can have a significant impact on inflation outcomes. Further, the historical record reveals that postwar periods can be disruptive, with sharp fluctuations in economic activity and inflation, and that a quick restoration of price stability requires a recalibration of monetary policy that cannot always be taken for granted.

23. The United States was not alone in having high inflation in the 1970s, but inflation experiences differed across countries. See, for example, Beyer et al. (2013), Ito (2013), or DiCecio and Nelson (2013).

DATA APPENDIX

The historical data used in Table 1 and Figures 1 and 2 are from www.measuringworth.com. These data have been used in many academic studies, some of which are cited in this article (e.g., Hall and Sargent 2021, 2022; and Rockoff, 2015).

The measures of the growth of the M2 money stock plotted in Figure 3 are derived from many sources. The source of pre-1959 money stock data is Carter et al. (2006, series cj28 for 1859-66; series cj45 for 1867-1947; series cj52 for 1947-59). For later years, the money stock is measured as M2 and sourced from the Board of Governors of the Federal Reserve System.

Data on the Fed's holdings of Treasury securities were graciously provided by George Hall. These data are monthly and measured in terms of market values. See George Hall, Jonathan Payne, Thomas J. Sargent, "US Federal Debt 1776-1960: Quantities and Prices," Working Papers 18-25, New York University, Leonard N. Stern School of Business, Department of Economics, 2018.

Data for Figures 5 through 8 are from the Bureau of Economic Analysis, the Federal Reserve System, and Haver Analytics. These are publicly available data that can be found in the St. Louis Fed's FRED® database.

APPENDIX: US WARTIME EXPERIENCES

The data indicate considerable variation in the extent and timing of money growth and inflation during major US wars and their postwar periods. This appendix provides a more detailed summary of the major wars from the Civil War to the Vietnam War for insights about the relationships between wartime government expenditures, inflation, and the monetary regime.

The Civil War

Civil War hostilities began with the Confederate attack on Fort Sumter in April 1861. Major hostilities ended on April 12, 1865, when General Robert E. Lee of the Confederate Army of Northern Virginia formally surrendered to US General Ulysses S. Grant; the war officially ended on June 2, 1865, when final terms of surrender were signed with the Confederate Army of the Trans-Mississippi.

Federal debt outstanding (excluding debt issued by the Confederacy) increased 15-fold during the Civil War, from about 2 percent of GDP in 1860 to 31 percent in 1865, where it remained until 1869. Over the same period, the money stock nearly tripled, from $554 million in 1860 to $1,445 million in 1865 (the peak estimate is $1,506 reached in November 1864) (Carter et al., 2006, series cj28). The consumer price index level rose from 8.1 to 15.8 (1982-84 = 100) and the GDP deflator rose from 4.7 to 8.4 (2012=100). Thus, while federal debt-to-GDP ratio increased by a factor of 15 during the war years, the money stock tripled and the price level roughly doubled. Although federal debt remained at about 31 percent of GDP in the immediate postwar years, the money stock and price level began to fall after the war ended.

The federal (Union) government was relatively successful in financing its war effort, such that the price level *only* doubled during the war, with an average annual inflation rate during 1861-65 of about 12 percent; but the Confederate government was much less successful in avoiding inflation. While tax revenue covered some 21 percent of federal expenditures, taxes paid for just 5 percent of spending by the Confederate government.[24] The Confederate government borrowed heavily to finance its expenditures and issued currency that could be used to purchase Confederate debt. Thus, in essence, the South financed much of its war effort by printing money. The result was an inflation rate that topped 9,000 percent by war's end.

24. See: http://www.taxhistory.org/www/website.nsf/Web/THM1861?OpenDocument.

The federal government had more ability to raise revenues through various forms of taxation. Tariff rates were increased on imported goods and new taxes imposed on incomes and a variety of domestic goods. The Internal Revenue Act of 1862 created the Bureau of Internal Revenue and provided for new excise taxes on goods and a variety of other taxes; a subsequent revenue act in 1864 raised tax rates and penalties for noncompliance (Studenski and Krooss 1952, pp. 137-60; Fox, 1986).

Although taxes covered a nontrivial portion of federal government wartime spending, a larger share was debt financed, including through noninterest-bearing debt issued in small denominations that was commonly referred to as "greenbacks." Greenbacks could be used to pay taxes, settle private debts, and purchase interest-bearing bonds authorized by the same legislation, but were not redeemable in gold or silver. A funding bill enacted in February 1862 authorized a $150 million greenback issue. Subsequent legislation in July 1862 and July 1863 authorized another $300 million of greenbacks. In total, the government issued $430 million of greenbacks, comprising approximately a third of the total money stock (Rockoff, 2015).

In addition to greenbacks, the federal government issued large amounts of interest- bearing debt. Congress helped spur demand for this debt by creating a new federal banking system and tying the currency issued by the banks in that system to the amount of government bonds they purchased. Before the Civil War, US currency comprised notes issued by state-chartered commercial banks. Although such notes were generally accepted in their local communities, notes of distant banks tended to circulate at discounts if they were accepted at all, reflecting uncertainty about whether the issuing bank would stand behind its promise to redeem them in gold or silver at par, as well as the cost of collecting from distant banks. The National Banking Acts of the 1860s standardized the currency by creating a new federal banking charter, requiring that banks chartered by the federal government ("national banks") hold government bonds to back their note issues and taxing notes issued by state-chartered banks out of existence (Studenski and Krooss, 1952, pp. 154-55). Along with greenbacks, national bank notes were the main form of currency in circulation in the United States before the founding of the Federal Reserve in 1914.

The Civil War era was one of fiscal dominance in the sense that the large expansion of the money stock was a direct result of the government's financing of its wartime expenditures. The money stock ceased to expand when the war ended and government spending fell. The United States did not have a central bank at the time to be dominated, however, and the money stock was determined largely by federal banking laws and the actions of the Treasury.[25] The bimetallic (gold and silver) standard was suspended during the war, which enabled the government to expand the money supply without the constraint of metallic backing for its note issues. There was little debate about returning to a metallic standard when the war ended; the only questions were about how and when to do so (Friedman and Schwartz, 1963, pp. 44). Although a subsequent period of slow economic growth and deflation led some to question resumption, the Resumption Act of 1875 mandated a return to a gold standard in 1879 at the pre-Civil War dollar-parity rate.[26] The United States remained on the gold standard until World War I, and growth of the money stock during 1879-1914 was determined largely by international gold flows and supply. The federal government typically ran budget surpluses in this era and fiscal policy had only limited impact on the monetary base (Friedman and Schwartz, 1963, Chapter 3). Figures 2 and 3 show that periods of monetary contraction and deflation were not uncommon.

25. The federal government chartered two proto-central banks (the first and second Bank of the United States) in the pre-Civil War era. The federal government owned 20 percent of the stock of each bank and they acted as fiscal agents for the government. However, their charters were not renewed; and, by the 1860s, the only banks operating in the United States were chartered by state governments. The United States had no official central banks until the Federal Reserve was founded in 1914.

26. Friedman and Schwartz (1963, pp. 44-50) describe the politics of returning to a metallic standard after the Civil War.

World War I

World War I broke out in Europe following the assassination of Austria's Archduke Franz Ferdinand on June 28, 1914. The United States initially adopted a policy of neutrality but German submarine aggression against neutral ships and ocean liners, such as the Lusitania, eventually pulled the United States into the conflict in early 1917. After some 18 more months of fighting, an armistice ending the conflict was signed on November 11, 1918.

Federal government expenditures increased sharply after the United States entered the war, from $713 million in 1916 to $1.95 billion in 1917, $12.7 billion in 1918, and $18.49 billion in 1919 (Carter et al., 2006, series ea585). Higher taxes covered a portion of the increase in spending. The outbreak of hostilities in 1914 reduced international trade and tariff revenue, which Congress attempted to offset with new excise taxes via the War Revenue Act of 1914. The Revenue Act of 1916 subsequently levied additional excise taxes, increased income taxes on households and corporations, and imposed an estate tax.[27] Additional legislation in 1917 levied a tax on excess corporate profits, raised personal income tax rates again, and added a surtax on household incomes above $6,000 (roughly $137,000 in 2022 dollars). Finally, the Revenue Act of 1918 (which was enacted in early 1919) raised tax rates on income, profits, and estates. The various changes to the federal tax system increased the share of federal revenues derived from income taxes from 16 percent in 1916 to as high as 58 percent during the years 1917-20.[28]

Despite substantial tax increases, the federal government also borrowed heavily to finance wartime spending. When the United States entered the war, the amount of federal government debt outstanding ($1.2 billion in 1916) was less than 5 percent of GDP. By 1918 it had risen to 30 percent and the amount of debt outstanding had ballooned to $12.5 billion. Total federal debt outstanding peaked in 1919 at $25.5 billion (Carter et al., 2006, series ea587).

Inflation began to rise before the United States entered the war. It peaked at over 20 percent and averaged above 10 percent per year from 1917 through 1920, despite government controls on the prices of various commodities (Stewart, 1941; Rockoff, 1984). In all, the price level more than doubled between 1914 and 1920, as did the stock of money.

The Federal Reserve System had been in operation less than three years when the United States entered World War I. The outbreak of war in 1914 precipitated a brief financial panic and a gold outflow from the United States. US exports began to increase within a few months, however, and gold poured into US banks, causing the money stock to rise.

Although the Fed played a major role in helping to finance the war effort, it purchased a relatively small portion of the government's debt. Figure 4 plots Federal Reserve holdings of marketable interest-bearing federal debt as a percentage of total federal interest-bearing debt outstanding from 1915 to 2021. Before the United States entered the war, the Fed held about 5 percent of outstanding Treasury debt. The percentage jumped to nearly 10 percent when the war began but quickly fell back to around 2 percent. While the United States was directly engaged in the war, the Fed's holdings of US government securities fluctuated around an average of $134 million. The Fed was actively engaged in both marketing and financing Treasury debt, however. The Reserve Banks were enlisted to assist in selling government securities to banks and the public and they encouraged their member banks to purchase Treasury securities by establishing preferential discount rates on loans secured by government notes and bonds (Meltzer, 2003, pp. 85-86; Studenski and Krooss, 1952, pp. 293-94). The Fed's loans to member banks increased from about $200 million in June 1917 to nearly $1.8 billion in November 1918 (Board of Governors, 1943, p. 373), and the vast majority were collateralized by US government securities (Board of Governors, 1943, p. 340). As of December 1918, the sum of the Fed's holdings of Treasury securities

27. The 16th Amendment to the U.S. Constitution, which permitted a federal income tax, was adopted in 1913 and Congress passed legislation authorizing an income tax that year.

28. See Studenski and Krooss (1952, pp. 280-301) for details on the financing of World War I.

and loans collateralized by Treasuries represented about 9 percent of total marketable Treasury debt outstanding.

The Fed's commitment to assisting the government's financing of wartime expenditures included support for debt issued after the war, such as the Victory Loan of 1919. Meltzer (2003, p. 90) argues that "by promising not to raise interest rates during the last wartime bond drive, the System relinquished a chance to moderate the postwar inflation." He also reports, however, that after the war, Fed officials were increasingly concerned about rising prices and began to press for higher interest rates, a proposal that the Treasury rejected. Thus, at least for a time, a strict regime of fiscal dominance prevailed in that monetary policy was dictated by the Treasury over the objections of Fed officials.

The period of fiscal dominance began to crack in 1919. Inflation continued to increase sharply, especially after the last of wartime price controls expired in mid-1919, and Reserve Bank officials renewed their calls for higher interest rates. In November, the Federal Reserve Board approved requests from several Reserve Banks to raise their discount rates. Continued inflation and a desire to protect the Reserve Banks' gold reserves motivated further rate hikes over the next six months (Meltzer 2003). The Treasury dropped its opposition to rate increases as government spending and the need to issue debt declined. By 1920, the federal budget was in surplus (Carter et al., 2006, series ea586) and the era of fiscal dominance ended.

Inflation fell sharply in the second half of 1920 and the economy entered a relatively protracted recession.[29] The GDP deflator fell 15 percent from 1920 to 1921, and consumer prices dropped 11 percent. Wartime budget deficits evaporated after the war, and the federal government ran modest annual surpluses throughout the 1920s. The money stock fell 5.6 percent on an annual average basis between 1920 and 1921, while monthly estimates show a drop of 9.4 percent from September 1920, when the money stock peaked, to January 1922.[30]

Thus, as in the Civil War era, World War I brought a large increase in federal spending, much of which was debt financed and supported by the Federal Reserve, resulting in a large increase in the money stock and inflation. The government budget deficit evaporated in 1920, which eliminated political pressure on the Fed to maintain low interest rates: Monetary policy tightened sharply, the price level fell, and the economy endured a moderately severe recession.

World War II

The inflation/deflation cycle that occurred after World War I was followed by nearly a decade of price stability. Then, significant deflation occurred during the first years of the Great Depression, 1929-33, when the price level fell some 30 percent. Federal debt increased relative to GDP: first because of a large decline in GDP from 1929 to 1933 and then from rising federal expenditures under the Roosevelt Administration's New Deal programs and increased defense spending in the second half of the decade.

Federal debt outstanding rose from 15 percent of GDP in 1929 to 42 percent in 1941. Annual federal government expenditures rose tenfold during World War II, from $9.5 billion in 1940 to $93 billion in 1945 (Carter et al., 2006, series ea680). A series of laws were passed that increased various excise taxes, raised tax rates, and broadened tax bases on corporate and personal incomes and inheritances. Notably, the Revenue Act of 1943 introduced withholding for the collection of personal income taxes.[31] Federal

29. The NBER dates the business cycle peak in January 1920 and the trough in July 1921.

30. The monthly money stock series referred to here is the sum of currency held by the public and commercial bank demand and time deposits, seasonally adjusted, as estimated by Friedman and Schwartz (1970) (NBER Macrohistory database series M14144: https://www.nber.org/research/data/nber-macrohistory-xiv-money-and-banking).

31. See Studenski and Krooss (1952, pp. 436-51) for information on federal government revenues and expenditures during World War II.

government revenues rose from $6.5 billion in 1940 to $45 billion in 1945 (Carter et al., 2006, series ea679). According to Hall and Sargent (2022), taxes financed 30.2 percent of war-related spending in World War II, interest-bearing debt financed 46 percent of the spending, and money growth financed 10.1 percent. The comparable numbers for World War I were 20.8 percent, 74.3 percent, and 6.9 percent, respectively.[32]

As it had during the first world war, the Federal Reserve played a major role in assisting the federal government in financing its debt in World War II. In World War I, the Fed's assistance took the form mainly of encouraging banks to purchase government bonds by providing them with preferential discount rates. By contrast, during World War II, the Fed purchased large amounts of government securities outright. Shortly after Pearl Harbor, the Fed announced that it would engage in open market operations to maintain specific yields on short- and intermediate-term Treasury securities and enforce a maximum market yield on long-term Treasury bonds. The Fed set the rate on 13-week Treasury bills at 3/8 percent, set the rate somewhat higher rates for longer-term securities, and enforced a maximum yield of 2.5 percent on long-term Treasury bonds. The policy required the Fed to purchase whatever securities the public did not want to hold at the specified yields. The result was an increase in the Fed's holdings of Treasury securities from $2.25 billion in December 1941 to $24.26 billion at the end of 1945 (Board of Governors of the Federal Reserve System, 1971, p. 468).

Given the pattern of rates set by the Fed and the public's expectation that the Fed would maintain the pattern at least for the duration of the war, the public preferred to invest in longer-term Treasuries rather than bills; and, by the end of 1945, the Fed held 75 percent of outstanding bills (Garbade, 2020). However, the Fed's holdings of Treasury bonds declined during the war to under $1 billion in August 1945 (Chaurushiya and Kuttner, 2003). As Table 1 shows, the Fed's holdings of Treasury debt ranged from about 5 percent of the total outstanding at the beginning of the war to 10 percent in 1945. The credibility of the Fed's commitment to the pegged yield levels likely encouraged the public to hold more Treasury debt than it would have otherwise. In addition, a scarcity of consumer goods and wartime rationing led households to save a higher percentage of their incomes than normal, and much of that savings was invested in government securities.

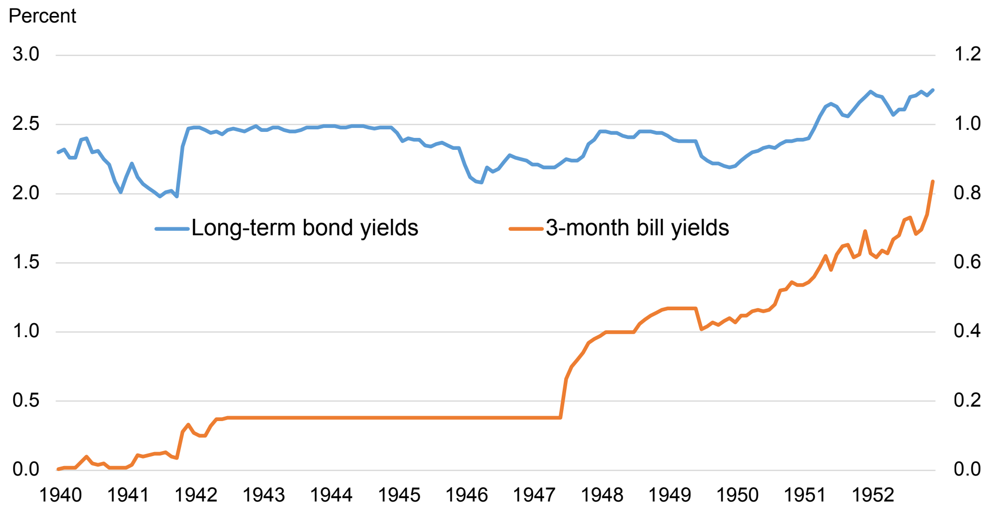
The 10-fold increase in the size of the Fed's Treasury portfolio during the war drove a large increase in the monetary base, which Friedman and Schwartz (1963) estimate rose from $23.7 billion in December 1941 to $43.3 billion in December 1945 (an 83 percent increase). Friedman and Schwartz (1970) estimate that the M2 money stock approximately doubled over the same period, from $64.1 billion to $132.7 billion.

The tremendous growth in federal expenditures and the money stock occurred when much of the productive capacity of the United States was devoted to producing war material. This likely would have resulted in significant inflation were it not for the imposition of price controls. The Emergency Price Control Act of January 1942 created the Office of Price Administration, which had the power to impose ceilings on all prices other than agricultural commodities and a system for rationing scare supplies of many goods. The controls imposed during World War II were "the longest and most comprehensive" in US history, according to Rockoff (1984, p. 85). In 1941, the CPI rose 5 percent, followed by a 10.7 percent increase in 1942 and a 6.1 percent increase in 1943. Thereafter inflation slowed sharply (see Figure 2). Controls were especially effective at containing inflation between April 1943 and June 1946, when strict enforcement limited the measured inflation rate to just 2.3 percent per year (Rockoff, 1984, p. 108).

32. Inflation and "other" contributors are the remaining sources of wartime financing. See Hall and Sargent (2022, Table 2).

Figure A1

Yields on Short- and Long-Term Treasury Yields, 1940-52



SOURCE: Federal Reserve and Haver Analytics.

The Postwar Period

Wartime government spending and growth in the money stock caused household demand to rise at a time when consumer goods were in short supply.

An interesting question is how much of the decline in the rate of inflation in 1948-49 was due to the increased supply of consumer goods and services rather than to a slowing of demand or a change in inflation expectations. Output of consumer goods expanded rapidly after the war, but fiscal and monetary policy also tightened. The last major government funding operation occurred in 1945. The fiscal deficit declined sharply in 1946, and the federal budget was in surplus during 1947-49 (Carter et al., 2006, series ea681).

Although the Fed remained committed to controlling yields on Treasury securities, the budget surpluses made that task easier: Yields remained at or below their pegged levels despite an inflation rate approaching 20 percent. At least partly in response to high inflation, the Fed ended its peg of the Treasury bill rate in July 1947. However, the pegs on other short- and intermediate-term securities, and the ceiling of 2.5 percent on the long-term bond, were maintained, and the Fed apparently had little difficulty keeping market yields from rising. Although prevented from making significant adjustments to interest rates by its commitment to keeping Treasury yields low and stable, the Fed increased reserve requirements on banks and reimposed wartime regulations on consumer installment credit in 1948 (Carlson and Wheelock, 2016). The growth rate of the money stock slowed and turned negative in early 1948, some nine months before the price level peaked in August 1948.

Following a mild recession in 1949, both the price level and money stock began to rise in early 1950. Treasury bill yields also began to rise, and the Fed made open market purchases to prevent yields on intermediate- and long-term Treasuries from moving above their pegs. Figure A1 illustrates the evolution of yields on 3-month Treasury bills and long-term government bonds. After the Fed stopped pegging the 3-month bill yield in July 1947, the market yield rose from 3/8 percent to 1 percent in early 1948. Although the long-term bond yield approached the Fed's 2.5 percent ceiling, it remained comfortably below the ceiling throughout most of 1946-49. However, by 1950 it had begun to rise along with the 3-month bill rate, despite increased Fed purchases. The outbreak of the Korean War in June 1950 brought increased defense expenditures and the US government ran a budget deficit in 1950.

Yields on Treasury securities continued to rise, which the Fed resisted with additional purchases. In May 1950, the size of the Fed's portfolio was at a low point. By March 1951, the Fed had added $5.5 billion to its Treasury holdings, a 32 percent increase, and inflation began to increase as well. The Fed found it increasingly difficult to prevent Treasury security yields from rising without greatly increasing its purchases. Facing a situation that it viewed as untenable, and with the support of key members of Congress, the Fed negotiated an "Accord" with the Truman Administration in March 1951 that ended its commitment to the interest rate pegs, thus ending the period of fiscal dominance.[33]

Friedman and Schwartz (1963, pp. 580-85) consider the contrast between 1947 and 1950-51: that is, respectively, between the period when the Fed was able to maintain its pegs on intermediate- and long-term bond yields with limited open market purchases, despite a 20 percent inflation rate and the period when it had more difficulty even though inflation was lower. They argue that inflation expectations were key: In 1946-47, the public expected that high inflation would be short-lived and foresaw an inflation/deflation pattern like that which followed World War I. Moreover, the public was still haunted by the specter of the Great Depression and the deflation of 1929-33 (Friedman and Schwartz, 1963, p. 658). Large federal government budget surpluses in 1947 and 1948 reinforced expectations of deflation and contributed to a slowing of demand pressures while helping the Fed maintain the structure of yields on Treasury securities with relatively little effort. But by July 1948, the Council of Economic Advisers was warning of "developing inflationary conditions which endanger both our domestic strength and our place in world affairs" (Economic Report of the President, p. 1). Conditions continued to deteriorate further. According to Friedman and Schwartz (1963, p. 610), the outbreak of the Korean War in June 1950 "drastically altered public expectations … and unleashed a speculative boom." Government spending and deficits rose sharply, inflation expectations became unanchored, and the Fed had to purchase large quantities of Treasury securities to prevent yields rates from rising above their pegged levels.

Korean War

The United States entered the Korean conflict in June 1950. Fighting continued until July 1953. Defense outlays boosted federal government expenditures and the federal budget swung from a surplus of $580 million in 1949 to a deficit of $3.1 billion in 1950. Legislation enacted in September 1950 raised corporate and personal income taxes, as well as various excise taxes, which put the budget back in surplus in 1951. Deficits returned in 1952 and 1953 with large spending increases, but taxes covered a higher percentage of wartime expenditures than in previous American wars. Although the stock of federal debt rose, GDP increased even more and the ratio of federal debt outstanding to GDP continued to fall (see Figure 1).[34]

As noted previously, beginning in 1950, the Federal Reserve made large purchases of Treasury securities to prevent their yields from rising above the levels agreed upon with the Treasury. Consequently, the money stock growth rate and inflation rose. The money stock had fallen at an average annual rate of 0.46 percent in 1949, but it began to increase in 1950 and averaged 4.1 percent between June 1950 and July 1953.

The decline in consumer prices ended when the United States entered the war. Inflation rose, peaking at 9.4 percent (measured on a year-over-year basis) in February 1951. The inflation rate then gradually declined. Friedman and Schwartz (1963, p. 598) argue that the Fed-Treasury Accord helped control inflation by reducing the liquidity of Treasury securities and inducing the public to hold larger

33. Meltzer (2003, pp. 691-712) describes the events and negotiations that led to the Accord.

34. Hall and Sargent (2021) note that federal debt outstanding had been falling before the war and contend that although the debt/GDP ratio declined during the war, it was higher in 1954 than it would have been had the war not occurred. Taking into account the halt in the paydown of the federal debt, Hall and Sargent estimate that 45.5 percent of Korean War expenditures were debt financed.

money balances and, perhaps more importantly, by limiting expected inflation. However, Rockoff (1984, pp. 185-87) argues that wage and price controls were more important than the Accord for reducing expected inflation, noting that inflation averaged just 2.1 percent while price controls were in place. In his view, (i) the increase in inflation that occurred early in the war was due more to inflationary expectations than to overly loose monetary policy and (ii) that the imposition of price controls in January 1951 lowered expected inflation. At the same time, the tax hikes and somewhat tighter monetary policy that followed the Accord prevented a burst of inflation after controls were lifted in early 1953. Rockoff (1984, p. 187) concludes that, by chance, "controls during the Korean War…were coordinated almost perfectly with monetary policy": The rate of inflation remained low after controls were lifted, averaging a mere 0.8 percent throughout 1953. Inflation remained low for the next dozen years, averaging just 1.4 percent from 1953 to 1965. Although the Fed refrained from making interest rate changes during periods when the Treasury was issuing new debt (a practice referred to as the "even keel" policy), the era of fiscal dominance of monetary policy ended with the Accord, at least for a while, and the percentage of outstanding marketable Treasury debt held by the Fed drifted downward.

Vietnam War

The "Vietnam era" began in February 1961 and ended in May 1975 according to the Congressional Research Service (2018). The US Department of Veterans Affairs, by contrast, dates the Vietnam war from 1964 to 1975. We use the latter dates since most of the casualties occurred from 1964 to 1975.[35]

The "Great Inflation" overlapped with the Vietnam War and occurred from 1965 to 1982 (Bordo and Orphanides, 2013). Over this period, the CPI increased by 74 percent. The start of the Great Inflation is widely attributed to sharply higher government spending, reflecting President Johnson's "guns and butter" policy—simultaneously prosecuting the war in Southeast Asia and the domestic "war on poverty," along with an accommodative monetary policy (e.g., Meltzer, 2005; Bernanke, 2022).

Figure A2 shows the growth in government spending during the Vietnam era measured relative to GDP. Three broad trends are evident. First, federal expenditures on defense rose between 1965 and 1967 but were generally declining over the longer period. Second, federal transfer payments increased sharply, from about 4.5 percent of GDP in the first quarter of 1966 to 10 percent of GDP in the third quarter of 1975.[36] Third, federal expenditures on nondefense goods and services, such as education or infrastructure, were roughly constant relative to GDP (a little more than 3 percent) from 1965 to 1975.

Despite increased government spending, and unlike the experience of prior major wars, the debt-to-GDP ratio declined between 1964 and 1975, from 39 percent to 25 percent, as inflation pushed nominal GDP higher while also eroding the real value of outstanding debt. However, the federal budget swung from a *surplus* equal to 0.3 percent of nominal GDP in fiscal year 1969 to a *deficit* of 4.1 percent of nominal GDP in 1976. Thus, the decline in the nominal value of Treasury debt relative to nominal GDP belied an expansionary fiscal policy.[37]

Total federal government spending continued to increase after 1975 in both nominal and real terms. Real defense spending peaked in 1968 and then fell until 1975 before rising again. However, nondefense spending was the primary driver of the growth in federal spending over the second half of the 1970s. This included the cost of servicing federal debt. Averaging less than 1.5 percent of GDP between 1962 and the mid-1970s, the growing federal debt and rising interest rates caused debt service costs to almost double to 2.6 percent of GDP by 1982.

---

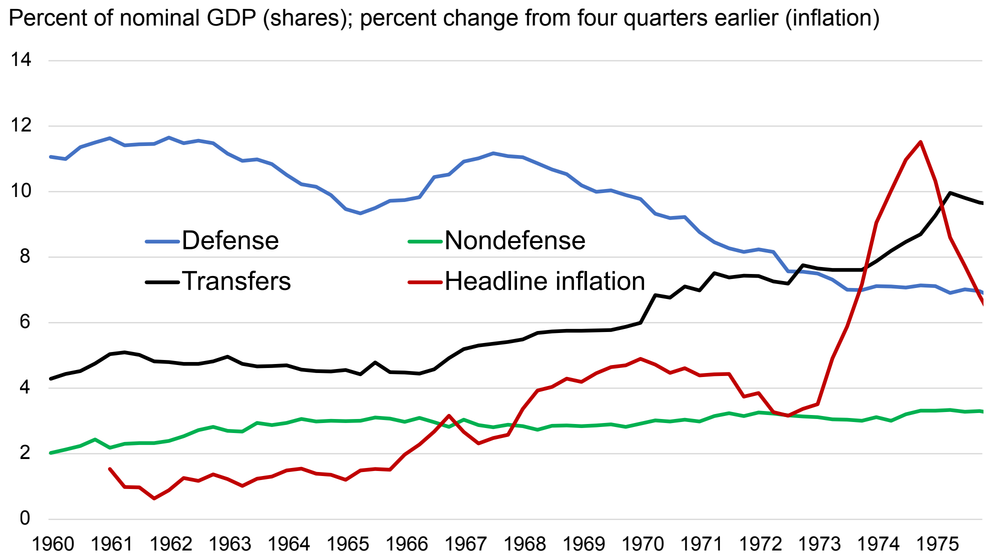
35. According to US government archives, there were 58,220 deaths during the war. There were 200 deaths from 1956 through 1963 (0.3% of the total). See https://www.archives.gov/research/military/vietnam-war/casualty-statistics.

36. Government transfer payments include social security benefits, medical benefits (e.g., from Medicare or Medicaid), veterans' benefits, and unemployment insurance benefits.

37. A falling debt-to-GDP ratio—both in nominal and real terms—largely reflected higher-than-expected inflation, which increased nominal GDP. See Neely (2022) for a discussion of these issues.

Figure A2

Government Expenditure Shares, Transfer Payments, and Inflation, 1960-75



SOURCE: BEA and Haver Analytics.

Unlike prior wars, inflation did not recede after the American withdrawal from Vietnam, but instead continued to rise to more than a 30-year high in 1980. The upward trajectory of inflation from the mid-1960s through the 1970s was punctuated by distinct surges, and when inflation declined following each surge the low point was invariably higher than the low that had followed the previous surge (see Figure 2). These surges and the rising trend rate of inflation have been widely blamed on the Fed's "stop-go" monetary policy (e.g., Meltzer 2009).[38]

Although the Fed would tighten policy when inflation was rising, it reversed course and eased substantially whenever unemployment rose and economic activity faltered. As evidence that monetary policy was to blame for the rising trend rate of inflation, many of the Fed's critics pointed to the rapid expansion of the money stock (see Figure 3), which more than doubled between 1964 and 1975 (rising at an average 8.1 percent annual rate). Money growth continued to accelerate after the Vietnam War, fueling the latter years of the Great Inflation. From 1976 to 1981, annual growth of the M2 money stock averaged 9.7 percent per year. Over this period, the CPI inflation rate averaged 9.2 percent per year, sharply higher than the average inflation rate of 4.9 percent that prevailed from 1964 to 1975.

The reasons for the Fed's failure to contain inflation during the late 1960s and 1970s have been widely studied and debated.[39] Much of the literature has focused on the policy views of the Fed's chairs at the time. The Federal Reserve had four chairs during the Great Inflation era—William McChesney Martin Jr., Arthur Burns, G. William Miller, and Paul Volcker.

Martin had generally orthodox views about the role of monetary policy and the causes of inflation. However, the Fed's monetary policies began to shift as the Kennedy and Johnson administrations appointed new members to the Fed's Board of Governors (and as the Fed hired staff economists with similar views). Policymakers came to accept the idea of a permanent tradeoff between unemployment and inflation, but also that unemployment could fall to very low levels (perhaps 4 percent or even lower)

38. See Blinder and Rudd (2013) for an alternative view that blames the inflation primarily on supply-side disturbances.

39. Among the many studies of Fed policy during the Great Inflation are Romer and Romer (2004), Meltzer (2005), Hetzel (2008) and Goodfriend and King (2013).

without triggering a substantial rise in inflation (Romer and Romer, 2004). As inflation began to accelerate under the Johnson Administration's "guns and butter" policy, Martin pressed President Johnson to enact an income tax surcharge in 1968 to help fight inflation. Then, after the 1968 election, he expressed to president-elect Nixon that "close relations and policy coordination [between the administration and the Fed] are important since monetary policy will need the help of fiscal policy to cope with inflation" (quoted by Bremner, 2004, p. 261).

Arthur Burns succeeded Martin as Fed chair in 1970. Like his predecessor, Burns believed that monetary policy should strive to achieve stable economic growth that was compatible with the aims of fiscal policy (Bremner, 2004, pp. 256-57). Burns also worked closely with the Nixon Administration on economic policy. According to Meltzer (2009, p. 760), Burns "sacrificed Federal Reserve independence and credibility for political reasons to a degree not seen since Marriner Eccles in the 1930s or perhaps never before." Burns strongly advocated the use of wage and price controls to contain inflation, persuaded a reluctant President Nixon to accept controls, and worked with the Administration to devise and administer the controls. Burns chaired the Committee on Interest and Dividends, which was charged with regulating administered interest rates such as the prime rate and dividends. The appointment set up a potential conflict of interest, which, according to Meltzer (2009, p. 767), might explain the Fed's reluctance to raise interest rates to combat inflation in 1972. Meltzer (2009, pp. 791-802) describes President Nixon's repeated efforts in 1971-72 to pressure Burns for a more expansionary monetary policy, reflected in faster growth of the money supply, and Burns' apparent acquiescence. However, Meltzer also notes that most of the other FOMC members and many academic economists supported a looser policy because the economy seemed to be growing at less than potential, as evidenced by an unemployment rate above 5 percent.[40]

Whereas Martin had comparatively orthodox views about the causes of inflation, Burns tended to blame inflation on non-monetary forces, including government budget deficits, oil price shocks and other supply factors, and monopolistic price setting by firms and labor unions. Further, Burns argued that the costs of using monetary policy to reduce inflation could be unacceptably high: "Once an inflation spiral gets underway, I am afraid there isn't a great deal that can be done constructively. A severe recession would bring it to a halt, but no one of us wants that. If we move back toward a policy of severe credit restriction, chances are that we will bring on a recession" (quoted by Bremner, 2004, p. 229).

G. William Miller succeeded Burns as Fed chair in 1978. Meltzer (2009, p. 910) contends that President Carter replaced Burns because he was viewed as "too much an anti-inflationist" and unwilling to cooperate with the Administration on economic policy. Carter's advisors thought that Miller would "work cooperatively with the Administration while preserving the independence of the Fed." (Meltzer, 2009, p. 923) Like Burns, Miller tended to blame inflation on supply-side factors, such as bad harvests and oil price shocks, as well as "excessive" increases in wage rates. Miller argued that monetary policy was not a particularly effective tool for reducing inflation and argued for policies that would increase competition and correct "structural problems" in the economy (Romer and Romer, 2004, pp. 142-43). Miller repeated Burns' mistake of never tightening policy sufficiently to bring inflation under control. Moreover, in Meltzer's (2009) view, the Fed lacked a coherent economic framework: Policymakers were too willing to blame inflation on supply shocks and rising wage rates, especially when the economy appeared to be growing at less than potential and unemployment was higher than what they viewed as consistent with full employment. Thus, unlike the early years of the Great Inflation era, when inflation stemmed at least in part from monetary accommodation of government spending, the continued rise of inflation in the second half of the 1970s seems largely driven by a lack of understanding among Fed

40. See also Woolley (1984, pp. 168-69) who notes that many Democratic members of Congress also favored faster money stock growth at the time.

officials and policymakers generally about the causes of inflation, their misreading of potential growth and employment, and fear that sufficient monetary tightening to control inflation would result in unacceptably high unemployment.

The Fed's monetary policy changed under Paul Volcker, who came into office in 1979 determined to restore price stability. Volcker publicly committed the Fed to controlling the growth of monetary aggregates, which he saw as necessary to establishing the Fed's anti-inflation credibility and the only way to persuade the FOMC to tighten policy sufficiently to break the inflation psychology that had become cemented in financial markets (Hetzel, 2008, pp. 152-53). The end of the Great Inflation thus occurred when Volcker and his colleagues on the FOMC implemented a monetary policy regime designed to slow the growth of money and bring down inflation expectations (Kliesen and Wheelock, 2021).

In the aftermath of the Great Inflation, inflation and macroeconomic volatility declined sharply, and the period from 1984 to 2006 is sometimes referred to as the Great Moderation.[41] Although there were occasional bursts of inflation, annual CPI inflation exceeded 5 percent only once during the period, when oil prices rose sharply following Iraq's invasion of Kuwait in 1990 (see Figure 2). Similarly, until 2020, annual M2 money growth never exceeded 10 percent. It was not until the COVID-19 pandemic that high and rising inflation and rapid money growth reemerged.

REFERENCES

Barro, Robert J. "On the Determination of the Public Debt," *Journal of Political Economy*, 1979, *87*(5), pp. 940-971.

Bernanke, Ben S. "The Great Moderation," Remarks at the Meetings of the Eastern Economic Association, Washington, D.C., February 20, 2004. https://www.federalreserve.gov/boarddocs/speeches/2004/20040220/.

Bernanke, Ben S. *21st Century Monetary Policy: The Federal Reserve from the Great Inflation to COVID-19*, W.W. Norton and Company, 2022.

Beyer, Adreas; Gaspar, Vitor; Gerberding, Christina; and Otmar Issing. "Opting Out of the Great Inflation: Germany Monetary Policy after the Breakdown of Bretton Woods," in Michael D. Bordo and Athanasios Orphanides, eds., *The Great Inflation: The Rebirth of Modern Central Banking*. Chicago: University of Chicago Press, 2013, pp. 301-56.

Bianchi, Francesco, and Leonardo Melosi. "Inflation as a Fiscal Limit," Paper prepared for the 2022 Federal Reserve Bank of Kansas City Economic Symposium, Jackson Hole, Wyoming, September 21, 2022. https://www.kansascityfed.org/Jackson%20Hole/documents/9121/Bianchi_JH2022.pdf.

Blinder, Alan S. and Jeremy B. Rudd. "The Supply-Shock Explanation of the Great Stagflation Revisited," in Michael D. Bordo and Athanasios Orphanides, eds., *The Great Inflation: The Rebirth of Modern Central Banking*. Chicago: University of Chicago Press, 2013, pp. 119-80.

Board of Governors of the Federal Reserve System. *Banking and Monetary Statistics 1914-41*. Washington, DC, 1943.

Board of Governors of the Federal Reserve System. *Banking and Monetary Statistics 1943-70*. Washington, DC, 1971.

Bordo, Michael D. "The Classical Gold Standard: Some Lessons for Today," Federal Reserve Bank of St. Louis *Review*, May 1981, pp. 2-17. https://files.stlouisfed.org/files/htdocs/publications/review/81/05/Classical_May1981.pdf.

Bordo, Michael D., and Mickey D. Levy. "Do Enlarged Fiscal Deficits Cause Inflation: The Historical Record," NBER Working Paper 28195, December 2020. https://www.nber.org/papers/w28195.

Bordo, Michael D., and Athanasios Orphanides, ed. *The Great Inflation: The Rebirth of Modern Central Banking*, The University of Chicago Press, 2013.

Borio, Claudio, Boris Hofmann, and Egon Zakrajsek "Does Money Growth Help Explain the Recent Surge in Inflation," *BIS Bulletin*, No. 67, January 26, 2023. https://www.bis.org/publ/bisbull67.htm.

Bremner, Robert P. *Chairman of the Fed: William McChesney Martin Jr. and the Creation of the American Financial System*, Yale University Press, 2004.

Bullard, James. "U.S. Macroeconomic Performance during the Pandemic with Three Topics for Future Research," Speech to the Czech National Bank Conference "Expectations in Dynamic Macroeconomic Models, August 30, 2021.

41. Bernanke (2004) discusses this period and cites the numerous studies and hypotheses that purport to explain this development. See https://www.federalreserve.gov/boarddocs/speeches/2004/20040220/.

https://www.stlouisfed.org/from-the-president/-/media/project/frbstl/stlouisfed/Files/PDFs/Bullard/remarks/2021/Bullard_Expectations_Dyn_Macro_Models_30_August_2021.pdf.

Carlson, Mark and David C. Wheelock. "Navigating Constraints: The Evolution of Federal Reserve Monetary Policy, 1935-1959," in M. Bordo and M. Wynne, eds., *The Federal Reserve's Role in the Global Economy: A Historical Perspective*. Cambridge: Cambridge University Press, 2016, pp. 50-83.

Carter, Susan B., Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright, eds. *Historical Statistics of the United States, Earliest Times to the Present: Millennial Edition*. New York: Cambridge University Press, 2006.

Chaurushiya, Radha and Ken Kuttner. "Targeting the Yield Curve: The Experience of the Federal Reserve, 1942-51." Board of Governors of the Federal Reserve System memorandum, June 18, 2003.

Cochrane, John H. "Fiscal Histories," *Journal of Economic Perspectives*, Vol. 36, No. 4, Fall 2022a, pp. 125-46.

Cochrane, John H. "Inflation Past, Present and Future: Fiscal Shocks, Fed Response, and Fiscal Limits," in Michael D. Bordo, John H. Cochrane, and John Taylor eds., *How Monetary Policy got Behind the Curve—and How to Get Back*. Stanford, CA: Hoover Institution Press 2022b, pp. 63-114.

Cochrane, John H. *The Fiscal Theory of the Price Level*, Princeton University Press 2022c.

Congressional Budget Office, "The Budget and Economic Outlook: 2022 to 2032," May 2022. https://www.cbo.gov/publication/58147.

Congressional Research Service, "U.S. Periods of War and Dates of Recent Conflicts," December 14, 2018. https://crsreports.congress.gov/product/pdf/RS/RS21405/28.

Congressional Research Service, "Tallying Federal Funding for COVID-19: In Brief," October 15, 2021. https://crsreports.congress.gov/product/pdf/R/R46449.

Council of Economic Advisers. *Midyear Economic Report of the President*, July 1948. https://fraser.stlouisfed.org/files/docs/publications/ERP/1948/ERP_1948_Midyear.pdf?utm_source=direct_download.

DiCecio, Riccardo and Edward Nelson. "The Great Inflation in the United States and the United Kingdom: Reconciling Policy Decisions and Data Outcomes," in Michael D. Bordo and Athanasios Orphanides, eds., *The Great Inflation: The Rebirth of Modern Central Banking*. Chicago: University of Chicago Press, 2013, pp. 393-448.

Federal Open Market Committee. "Summary of Economic Projections," Various. https://www.federalreserve.gov/monetarypolicy/fomccalendars.htm.

Fox, Cynthia G. "Income Tax Records of the Civil War Years," National Archives, *Prologue Magazine*, Winter 1986, *18*(4), Genealogy Notes. https://www.archives.gov/publications/prologue/1986/winter/civil-war-tax-records.html.

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

Friedman, Milton and Anna J. Schwartz. *Monetary Statistics of the United States*. New York: Columbia University Press, 1970.

Garbade, Kenneth, "Managing the Treasury Yield Curve in the 1940s." Federal Reserve Bank of New York Staff Report No. 913, February 2020.

Giovanni, Julian di. "How Much Did Supply Constraints Boost U.S. Inflation?" Federal Reserve Bank of New York Liberty Street Economics, August 24, 2022. https://libertystreeteconomics.newyorkfed.org/2022/08/how-much-did-supply-constraints-boost-u-s-inflation/.

Goodfriend, Marvin, and Robert G. King. "The Great Inflation Drift," in Bordo and Orphanides, ed., *The Great Inflation: The Rebirth of Modern Central Banking*, The University of C Chicago Press, 2013, pp. 181-209.

Hall, George J., and Thomas J. Sargent. "Debt and Taxes in Eight U.S. Wars and Two Insurrections," *Handbook of Historical Economics*, Bisin and Federico, eds., Elevier Inc., 2021, pp. 825-80.

Hall, George J. and Thomas J. Sargent. "Three World Wars: Fiscal-Monetary Consequences." PNAS 119, no. 18, 2022.

Hetzel, Robert L. *The Monetary Policy of the Federal Reserve: A History*. Cambridge University Press, 2008; https://doi.org/10.1017/CBO9780511754173.

Ito, Takashito. "Great Inflation and Central Bank Independence in Japan," in Michael D. Bordo and Athanasios Orphanides, eds., *The Great Inflation: The Rebirth of Modern Central Banking*. Chicago: University of Chicago Press, 2013, pp. 357-92.

Keen, Michael, and Joel Slemrod. *A Rebellion, Rascals, and Revenue: Tax Follies and Wisdom Through the Ages*. Princeton University Press, 2021.

Kindleberger, Charles P. *A Financial History of Western Europe*. London: George Allen & Unwin, 1984.

Kliesen, Kevin L. and David C. Wheelock. "Managing a New Policy Framework: Paul Volcker, the St. Louis Fed, and the 1979-82 War on Inflation." Federal Reserve Bank of St. Louis *Review*, First Quarter 2021, *103*(1), pp. 71-97.

Lucas, Robert E., and Nancy L. Stokey. "Optimal Fiscal and Monetary Policy in an Economy Without Capital," *Journal of Monetary Economics*, Vol. *12*(1) 1983, pp. 55-93.

Martin, Fernando M.. "Fiscal Dominance," Federal Reserve Bank of St. Louis Working Paper 2020-040B, September 2021. https://s3.amazonaws.com/real.stlouisfed.org/wp/2020/2020-040.pdf.

Meltzer, Allan H. *A History of the Federal Reserve, Volume 1, 1913-1951*. Chicago: University of Chicago Press, 2003.

Meltzer, Allan H. "Origins of the Great Inflation," Federal Reserve Bank of St. Louis *Review*, March/April 2005, vol. *87*(2, Part 2), pp. 145-75. https://files.stlouisfed.org/files/htdocs/publications/review/05/03/part2/Meltzer.pdf.

Meltzer, Allan H. *A History of the Federal Reserve, Volume 2, Book 1 and Book 2, 1970-1986*. Chicago: University of Chicago Press, 2009; https://doi.org/10.7208/chicago/9780226519968.001.0001.

Neely, Christopher J. "Inflation and the Real Value of Debt: A Double-Edged Sword," Federal Reserve Bank of St. Louis *On The Economy Blog*, August 1, 2022. https://www.stlouisfed.org/on-the-economy/2022/aug/inflation-real-value-debt-double-edged-sword.

Officer, Lawrence H. and Samuel H. Williamson, *MeasuringWort*h, 2022 (www.measuringworth.com).

Ohanian, Lee E. "The Macroeconomic Effects of War Finance in the United States: World War II and the Korean War," *American Economic Review*, March 1997, Vol. *87*(1), pp. 23-40.

Poast, Paul. "Central Banks at War," *International Organization,* Winter 2015, *69*(1), pp. 63-95.

Powell, Jerome H. "Monetary Policy in the Time of COVID," Remarks at the "Macroeconomic Policy in an Uneven Economy", economic policy symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming (via webcast), August 27, 2021. https://www.federalreserve.gov/newsevents/speech/powell20210827a.htm.

Rockoff, Hugh. "Price and Wage Controls in Four Wartime Periods," *The Journal of Economic History*, Vol *41*(2), June 1981.

Rockoff, Hugh. *Drastic Measures: A History of Wage and Price Controls in the United States*. Cambridge: Cambridge University Press, 1984.

Rockoff, Hugh. "War and Inflation in the United States from the Revolution to the First Iraq War," National Bureau of Economic Research Working Paper 21221, May 2015.

Romer, Christina D. and David H. Romer. "Choosing the Federal Reserve Chair: Lessons from History," *Journal of Economic Perspectives*, Winter 2004, *18*(1), pp. 129-62.

Rouse, Cecilia, Jeffrey Zhang, and Ernie Tedeschi. "Historical Parallels to Today's Inflationary Episode," U.S. Council of Economic Advisers Blog, July 6, 2021. https://www.whitehouse.gov/cea/written-materials/2021/07/06/historical-parallels-to-todays-inflationary-episode/.

Sargent, Thomas J. "The Ends of Four Big Inflations," in *Inflation: Causes and Effects*, Robert E. Hall (ed.), National Bureau of Economic Research, University of Chicago Press, 1982, pp. 41-97.

Schnabel, Isabel. "The Shadow of Fiscal Dominance: Misconceptions, Perceptions, and Perspectives," Speech at the Centre for European Reforms and the Eurofi Financial Forum on "is the Current ECB Monetary Policy Doing More Harm than Good and What are the Alternatives, Berlin, September 11, 2020. https://www.ecb.europa.eu/press/key/date/2020/html/ecb.sp200911~ea32bd8bb3.en.html.

Shapiro, Adam Hale. "How Much Did Supply and Demand Drive Inflation," Federal Reserve Bank of San Francisco *Economic Letter*, June 21, 2022. https://www.frbsf.org/wp-content/uploads/sites/4/el2022-15.pdf.

Stewart, Stella. "Government Price Control in the First World War," *Monthly Labor Review*, February 1941, *52*(2), pp. 271-85.

Studenski, Paul and Herman E. Krooss. *Financial History of the United States: Fiscal, Monetary, Banking and Tariff, including Financial Administration and State and Local Finance*. New York: McGraw-Hill, 1952.

Summers, Lawrence H. "On Inflation, We Can Learn From the Mistakes of the Past—Or Repeat Them," *The Washington Post*, February 3, 2022.

Timiraos, Nick. "Can Central Banks Maintain their Autonomy?" *The Wall Street Journal*, August 26, 2022.

Volcker, Paul A. Interview. *Federal Reserve Board Oral History Project*. Transcript, 2008-10:

Waller, Christopher J. "Reflections on Monetary Policy in 2021," Remarks at the 2022 Hoover Institution Monetary Conference, Stanford, California, May 6, 2022. https://www.federalreserve.gov/newsevents/speech/waller20220506a.htm.

Williamson, Stephen. "High Inflation: Where Did it Come From" (Part 1) and "High Inflation: What to Do About It?" June 23, 2022. Part I: https://stevewilliamson.substack.com/p/part-i-high-inflation-where-did-it. Part II: https://stevewilliamson.substack.com/p/part-ii-high-inflation-where-did.

Woolley, John T. *Monetary Politics: The Federal Reserve and the Politics of Monetary Policy*. Cambridge: Cambridge University Press, 1984.

Demographic Disparities in COVID-19 Disruptions: What Has Shaped Them?

Violeta A. Gutkowski

Abstract

This article uses the Community Impact Survey implemented by the Federal Reserve System in 2021 to identify COVID-19 disruptions on low- to moderate-income communities. I find that communities that were primarily of Color were more likely to be significantly disrupted by COVID-19 than White communities. I also assess the importance of certain challenges, such as returning to work or unequal access to government relief, in shaping the observed demographic differences in disruption.

JEL codes: D04, D63

Federal Reserve Bank of St. Louis *Review*, Fourth Quarter 2023, *105*(4), pp. 261-79.
https://doi.org/10.20955/r.105.261-79

1. INTRODUCTION

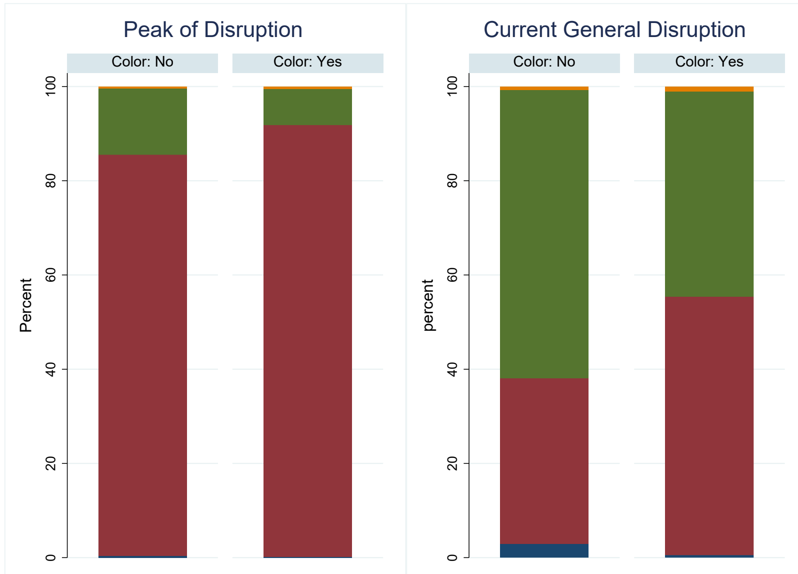
In the past years, we have seen waves of COVID-19 infections disrupting communities in the United States. The pandemic has interfered with lives across the country, and its effects on low- to moderate-income (LMI) individuals and communities have been significant. Though the peak of distress has passed, people of Color are still bearing major disturbances. Figure 1 shows that at the peak of COVID-19, distressed communities were facing serious disruptions regardless of demographic composition. However, 18 months into the pandemic, communities of Color were 20 percentage points more likely to be significantly disrupted than White communities even though the disruptions had already halved.[1]

This article analyzes a different and unique channel to assess the disruption gap between the two communities—unequal access to government COVID 19 relief—and compares it with other possible explanations, such as challenges returning to work. Government relief has been very critical to LMI communities (Chalise and Gutkowski, 2021). It has kept millions out of poverty, helped people access health coverage, and reduced hardships like the inability to afford food or meet other basic needs (Budget and Policy Priorities, 2021). Nevertheless, equally distributing these funds was challenging. Holtzblatt and Karpman, 2020 show that the receipt of Economic Impact Payments from the CARES Act occurred more slowly—or not at all—for some groups. Having no recent history of filing taxes, lacking internet access, and being unbanked were some of the barriers to receiving government payments. Many were also excluded from unemployment benefits. For example, people of Color who often work in low-wage or part-time jobs, or as independent contractors, may have been more likely to be excluded or not eligible for regular unemployment insurance (Janger, Rubin, and Singh, 2020).

1. The distinction between primarily White communities and communities of Color comes from the following survey question: "Does the entity you represent primarily serve a community of Color?" See Section ??.

Violeta A. Gutkowski is a lead analyst in the Institute for Economic Equity at the Federal Reserve Bank of St. Louis. The author thanks William Rodgers and Nishesh Chalise for their suggestions and support.

Figure 1
General Economic Disruption



NOTE: The figure shows the percentage of respondents who reported disruption at the peak of distress (left panel) and at the time of the survey (right panel) for entities serving communities that were primarily of Color and those that were not.*

*Questions on disruption were the following: "Thinking about the period of peak of distress caused by the pandemic, what level of disruption did COVID-19 have on economic conditions in your community at that time?" "Currently, what level of disruption is COVID-19 having on economic conditions in your community?"

Furthermore, not all small businesses had access to Paycheck Protection Program (PPP) loans due to lack of eligibility, burdensome paperwork, and lack of knowledge and understanding of the program. According to Liu and Parilla, 2020, it took 31 days for small businesses with paid employees in majority-Black zip codes to receive PPP loans—7 days longer than those in majority-White communities. Lack of a bank relationship was also a burden. The initial demand for loans significantly exceeded the ability of banks to deliver them, and they appear to have prioritized firms with which they had a preexisting relationship (Granja et al., 2022; Li, Strahan, and Zhang, 2020). Thus, larger firms—which tend to have ongoing bank relationships—accessed PPP funds sooner than smaller firms, on average.

I find that unequal access to government relief plans was as important as challenges in returning to work in explaining variation in disruption between White communities and communities of Color. Returning to work has been particularly challenging to many due to risk of exposure, lack of childcare, or lack of public transportation. Together, these challenges can explain approximately 30 percent of the disruption gap. These findings highlight the importance of targeting, implementing, and distributing government relief packages and the relevance of childcare in the US economy.

The difference in the disruption across demographics groups could be explained by a variety of factors. For example, Maxwell, 2020 suggests that economic hardship, food and housing insecurity, and unequal access to mental healthcare services could have contributed to explaining differences in the overall disruption caused by COVID-19. Gemelas, Davison, et al., 2022 show that people of Color experienced larger income losses, which could have made it harder for them to make ends meet. In addition, workers in communities of Color had less ability to telecommute, which may have resulted in unequal health risks affecting overall employment and labor force participation (OECD, 2022).

Differences in access to healthcare and health insurance could also have contributed to the gap. Gangopadhyaya, Karpman, and Aarons, 2020 show that between late April and mid-July, more than three million adults lost employer-sponsored insurance (ESI). These losses were concentrated among Hispanic adults, young adults, men, and adults who did not attend college. Though the reduction in ESI was partially offset by a rise in public coverage, groups that faced the largest ESI losses also saw the largest increases in uninsurance.

The next section provides more information about the data from the Community Impact Survey (CIS) as well as the main variables of interest. Section 3 presents the main analysis and results, followed by some robustness exercises. I conclude in Section 4 with some discussion about the findings and further possible areas of research toward an equitable recovery path.

2. DATA AND ANALYSIS

2.1 Community Impact Survey (CIS)

The data for this analysis come from the CIS, conducted by the Federal Reserve System in collaboration with eight national partners in August 2021.[2] The survey was designed to monitor the impact of the pandemic on low-to moderate-income (LMI) communities and the organizations that serve them. It provides an insightful and informative snapshot of how COVID-19 affected people and organizations as of the dates the survey was administered.

Survey respondents included representatives of nonprofit organizations, government agencies, financial institutions, and other organizations who serve LMI communities across the US.[3] Responses were collected through a convenience sampling method that relied on contact databases to identify representatives of these community organizations. These representatives were invited to participate in the survey via emails, newsletters, and social media posts.

The survey was organized into three different sections: respondent profile questions, questions about the impact of COVID-19 on the entities providing services, and questions about the impact on the people and communities these entities served. It is worth noting that respondents to this survey were service provider organizations, and thus all of the answers about the impact on people and communities come from their perspective.

This article focuses on the disruption caused by COVID-19 on LMI people and communities, differences in the impact across demographic groups, and how challenges returning to work or accessing government funding have shaped the observed differences in disruption. Given the diversity of responding entities, I focus on nonprofit organizations that completed the full survey (N=2291). In the next section, I describe the main variables of interest for this work.

2.1.1 Communities of Color

Respondents were asked if they served a community that was primarily of Color (possible responses were yes, no, or unsure). Throughout the article, I refer to "community of Color" as those who responded with yes (47 percent of responses) to this question and "White" as those who responded with no (48 percent of responses).[4,5] Communities that were primarily of Color included neighborhoods where the primary population was American Indian or Native Alaskan (4 percent), Asian (3 percent), Hawaiian or Pacific Islander (2 percent), Hispanic or Latino (41 percent), non-Hispanic Black or African American (45 percent), and Other (5 percent).

2.1.2 Disruption

My main variable of interest is the COVID-19 level of economic disruption on people and communities. The survey question asked "Currently, what level of disruption is COVID-19 having on economic conditions in your community?" (possible responses were significant disruption, some disruption, no disruption, or unsure). Figure 1 shows the general economic disruption on LMI communities disaggregated by the demographics served. At the peak of distress, disruption levels were high for all demographic groups, with more than 80 percent of respondents reporting significant disruption. While significant disruption almost halved in 2021 for all LMI communities, it was 20 percentage points more likely for communities of Color to be significantly disrupted than White communities.

While economic conditions could be interpreted differently across respondents, they provide a broad overview of how respondents saw COVID-19 disruptions impacting their communities. To have a better understanding of COVID-19's impact on several dimensions, the survey also asked about disruption in different segments of the economy that are vital for the development of LMI communities. These segments were financial stability (related to income loss and income instability), small business (including short-/long-term closure, supply chain disruptions, and reduced demand), access to healthcare (such as access to adequate healthcare, access to health insurance, and mental health services), services for children (including the availability of childcare and adequate access to K-12 education), housing stability (involving evictions, back rent, foreclosures, and homelessness), and basic consumption needs (for example, food, household essentials, and other personal needs).

---

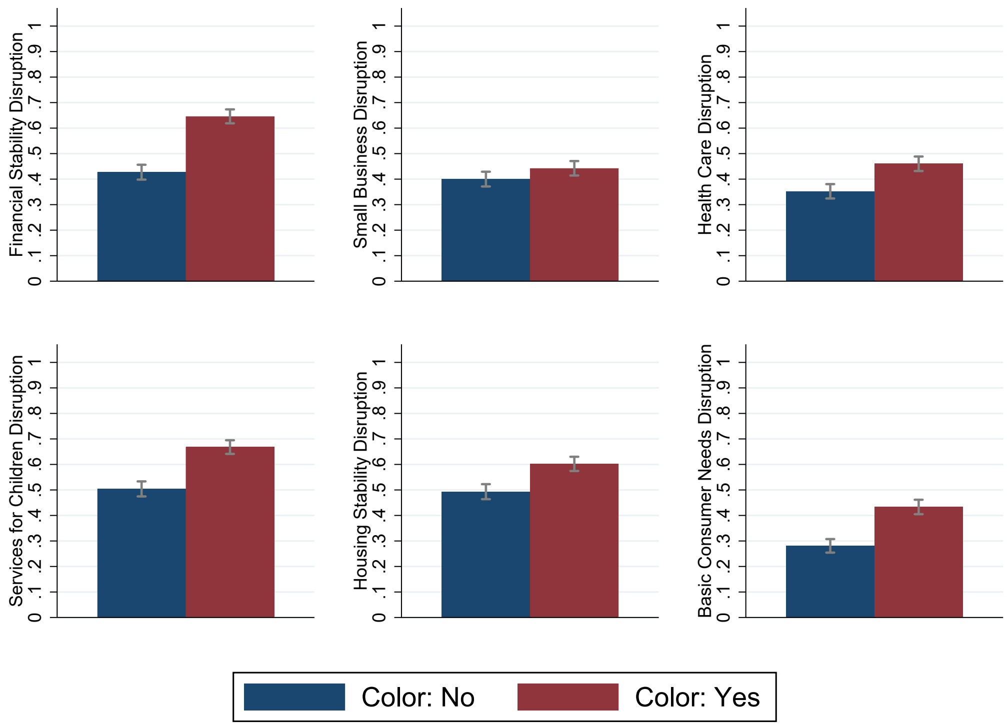
2. For more information on the survey, please visit https://fedcommunities.org/data/main-street-covid19-survey-2021.

3. There were approximately 3,700 responses from entities who answered most of the survey, including the questions related to being a direct service provider or not and COVID-19's level of disruption during the peak of the pandemic and at the time of the survey.

4. Five percent of responses were "unsure," which are not included in the study.

5. The survey did not explicitly ask whether the respondent served a primarily White community.

Figure 2
Significant Disruption in Segments of the Economy



NOTE: The figure shows the percentage of respondents who reported significant disruption in their communities in each segment of the economy at the time of the survey, by the demographics served with 95 percent confidence interval bars.

There was some variation in responses for these different segments of the economy with significant disruption ranging between 30 and 60 percent of responses (Appendix Figure A.3). Nevertheless, communities of Color were more likely to be significantly disrupted than White communities for all segments of the economy surveyed (Figure 2). Appendix Tables A.1 and A.2 provide a disaggregation of responses by demographic composition, type of engagement with community and location, and the distribution of impact for each segment.

2.1.3 Challenges

The next set of relevant variables relate to the barriers or challenges that affected people and communities.[6] Respondents were asked to rate the seriousness of the following challenges: 1) returning to work (such as childcare, public transportation, COVID-19 exposure risk), 2) accessing government funding (lack of eligibility or capacity for processing applications), 3) relationship with banks to access capital, 4) and applying for funds (including complex process, burdensome paperwork, internet access).

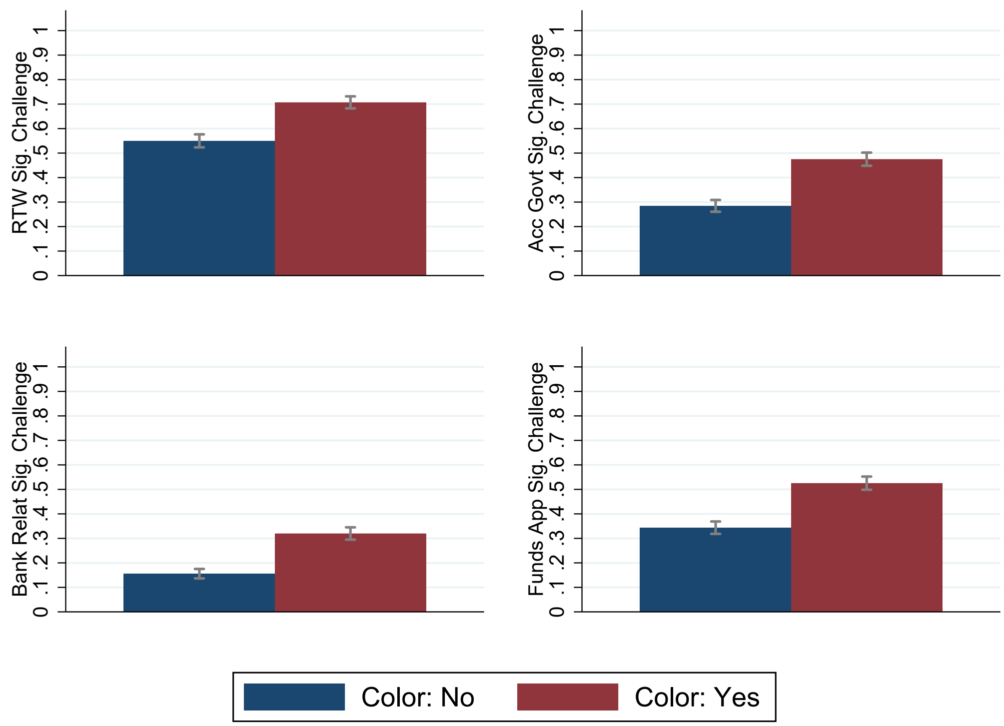
More than 65 percent of responses indicated that returning to work was a significant challenge and 38 percent claimed that accessing government funding was a severe hurdle. Lack of a bank relationship was a serious barrier for almost 30 percent of respondents, and more than 45 percent indicated applying for funds was a severe obstacle. Appendix Table A.3 provides the distribution of challenges among respondents. However, this distribution was not symmetric across demographic groups: The likelihood of these obstacles being severe was higher for communities of Color.[7] Figure 3 shows there is a statistically significant difference in the likelihood of communities of Color reporting that these difficulties were acute relative to White communities."

Differences in barriers to returning to work could be related to unequal access to childcare due to high costs, inflexible schedules, closures, and illnesses due to COVID-19 exposure. Lee and Parolin, 2021 find that Black, Latino, and Asian families have been exposed to childcare closures to a much greater extent than White families. Labor market segmentation and differences in occupations could also account for the observed differences in

6. They were asked the following question: "During the COVID-19 pandemic, how challenging were the following for the people and communities you serve?" Respondents replied with "significant challenge," "challenge," "no challenge," or "unsure."

7. See Appendix Figure A.4 for responses across White communities and communities of Color.

Figure 3
Significant Challenges by Demographic Group



NOTE: The figures show the percentage of respondents who reported significant challenges for the communities they serve by demographic group served. Top left: Return-to-work challenges (such as childcare, public transportation, COVID-19 exposure risk). Top right: Challenges accessing government funding (lack of eligibility or capacity for processing applications). Bottom left: Relationship with banks to access capital. Bottom right: Applying for funds (including complex process, burdensome paperwork, internet access).

difficulties returning to work. Accessing government funds has been a challenge to many, particular to people of Color due to lack of eligibility, lack of knowledge or resources to apply to them, and lack of historical bank relationships, which favored previous connections when expediting funds. Holtzblatt and Karpman, 2020 show that the receipt of Economic Impact Payments from the CARES Act occurred more slowly—or not at all—for some groups. Barriers to payment receipt included having no recent history of filing taxes, lacking internet access, and being unbanked. Autor et al., 2022 show that the PPP—another well-known federal program to mitigate COVID-19 disruptions—incidence was highly regressive, with about three-quarters of PPP funds accruing to the top quintile of households.

2.1.4 Control Variables

Within the respondent profile part of the survey, entities were asked whether they served primarily in urban, suburban, or rural communities and whether or not they were offering services directly to people or small businesses (direct service providers). Suburban observations account for 13 percent of observations with a similar demographic composition and outcome than rural areas. I exclude these observations to reduce the amount of noise introduced in the analysis and to facilitate the interpretation of the coefficient on location. Given that the geographic location of service as well as direct service (or not) could have a different demographic composition, I control for these variables in my analysis (see Appendix Figures A.1 and A.2).

3. RESULTS

In this section, I first provide supportive evidence that communities that were facing major challenges were more likely to be significantly disrupted at the time of the survey. I then show, in Section 3.2, that the disruption gap between White communities and communities of Color was approximately 17 percentage points and that incorporating these challenges reduces the disruption gap by a third. Section 3.2.1 assesses the relative importance of each obstacle in explaining the disruption gap. The main finding is that all the challenges

Table 1
Challenges and Likelihood of Disruption

	General economic significant disruption					
	(1)	(2)	(3)	(4)	(5)	(6)
Urban	0.124***	0.108***	0.0973***	0.105***	0.0997***	0.0837***
	(0.0215)	(0.0212)	(0.0212)	(0.0214)	(0.0213)	(0.0210)
DSP	0.0474	0.0489*	0.0443	0.0455	0.0375	0.0420
	(0.0293)	(0.0288)	(0.0283)	(0.0286)	(0.0284)	(0.0279)
RTW		0.198***				0.116***
		(0.0209)				(0.0224)
Acc Gvt Fund			0.238***			0.137***
			(0.0209)			(0.0253)
BK Relat				0.211***		0.0716**
				(0.0238)		(0.0280)
Fund Appl					0.212***	0.0813***
					(0.0207)	(0.0260)
Constant	0.342***	0.225***	0.270***	0.305***	0.273***	0.192***
	(0.0302)	(0.0315)	(0.0294)	(0.0296)	(0.0297)	(0.0306)
Observations	2,291	2,291	2,291	2,291	2,291	2,291
R-squared	0.015	0.051	0.069	0.048	0.059	0.095

NOTE: Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

have been similarly important in explaining the disruption gap. Section 3.3 provides results for my preferred specification and contrasts challenges returning to work with overall access to government support during the pandemic. Last, in Section 3.4 I perform some robustness checks on the impact on a variety of segments of the economy to further support the main findings as well as to control for median household income in the environment (Section 3.5).

3.1 Challenges and Likelihood of Disruption

I first provide evidence of the positive correlation between the challenges and the likelihood of being significantly disrupted. I use the following regression:

$$SigDisrup = \alpha + \beta * Challenge + \gamma * urban + \delta * DSP + \upsilon, \quad (1)$$

where *SigDisrup* is a dummy variable that equals one if disruption was significant and equals zero otherwise. *Challenge* is a dummy variable that equals one if the respondent indicated significant challenge and equals zero otherwise. The four challenges analyzed are returning to work (*RTW*), accessing government funding (*Acc Gvt Fund*), bank relationship (*BK Relat*), and the fund application process (*Fund Appl*). To control for differences in disruption due to geographic location, I incorporate *Urban*, a dummy variable that equals one if the entity offered services primarily in urban areas. *DSP* is a dummy that equals one if the responding entity was a direct service provider (offering services directly to individuals, families, or small business owners). Controlling for the type of engagement with the community is important since direct service providers could be less optimistic about disruption and recovery than indirect providers given the nature of their engagement with the community, and thus the results could be biased (see Appendix Table A.2).

Table 1 shows that the challenges considered are relevant in explaining the variation in the disruption as well as the location of service. The communities that were more likely to face any of these challenges were more likely to be significantly disrupted given the positive sign and statistical significance of the coefficients. The table shows that urban areas were more likely to be severely disrupted. Consistent with earlier findings by Brooks, Mueller, and Thiede, 2021 and Parker, Horowitz, and Minkin, 2021, this could be due to higher exposure risk in urban areas than rural areas as well as distinct disruptions in the labor market. Column 2

Table 2
Demographics, Challenges, and COVID-19 Economic Disruption

	General economic significant disruption						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Color	0.191***	0.169***	0.143***	0.131***	0.140***	0.137***	0.108***
	(0.0205)	(0.0219)	(0.0219)	(0.0220)	(0.0223)	(0.0220)	(0.0220)
RTW			0.178***				0.108***
			(0.0211)				(0.0224)
Acc Gvt Fund				0.217***			0.129***
				(0.0212)			(0.0252)
BK Relat					0.184***		0.0594**
					(0.0240)		(0.0278)
Fund Appl						0.191***	0.0765***
						(0.0209)	(0.0259)
Urban		0.0649***	0.0599***	0.0542**	0.0587***	0.0545**	0.0493**
		(0.0227)	(0.0222)	(0.0223)	(0.0226)	(0.0223)	(0.0220)
DSP		0.0333	0.0368	0.0336	0.0340	0.0270	0.0333
		(0.0290)	(0.0286)	(0.0281)	(0.0285)	(0.0283)	(0.0278)
Constant	0.365***	0.306***	0.205***	0.248***	0.280***	0.250***	0.180***
	(0.0145)	(0.0300)	(0.0313)	(0.0293)	(0.0295)	(0.0296)	(0.0305)
Observations	2,291	2,291	2,291	2,291	2,291	2,291	2,291
R-squared	0.037	0.040	0.069	0.084	0.064	0.075	0.105

NOTE: Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

shows that communities that faced serious barriers in returning to work were almost 20 percent more likely to be heavily distressed at the time of the survey. Lack of regular hours for childcare, unpredictable closures due to COVID-19 exposure or illnesses, and inflexible schedules all could have contributed to these observed differences.

Similarly, communities with major issues accessing government funds were also more likely to be seriously disrupted (Column 3). Federal COVID-19 relief has helped millions of people navigate the crisis and has kept many out of poverty. However, those who were not able to access these funds could not benefit from the generous package. In addition, communities that lacked a bank relationship (Column 4), as well as those that faced burdensome paperwork, lack of internet, and a complex application process (Column 5), were approximately 20 percent more likely to be extremely disrupted by COVID-19. These barriers related to being unbanked, not being within the network of lending banks, lacking knowledge, or lacking capacity were clear obstacles in the process of applying and thus accessing funds. Column 6 incorporates all four challenges together. The coefficients on each challenge fall; however, all four challenges remain positive and significantly different from zero, indicating their relevance when assessing the disruption.

3.2 *Challenges and the Disruption Gap*

In this section I focus on the disruption gap between White communities and communities of Color and find that the explored challenges can explain almost 30 percent of the disruption gap. To this end, I run a similar regression as before but now include *Color*—a dummy variable that equals one if the community served was primarily of Color—as one of my main explanatory variables of interest:

$$(2) \qquad SigDisrup = \alpha + \theta * Color + \beta * Challenge + \gamma * Urban + \delta * DSP + \upsilon.$$

Table 3 shows the results. The first two columns indicate that communities of Color were 19 percentage points more likely to experience significant disruption at the time of the survey and almost 17 percentage points when accounting for differences in location and type of engagement with the community. Columns 3 to 6 incorporate, one by one, the different challenges of interest. In all cases, the coefficients on the challenges

Table 3
Demographics, Challenges, and COVID-19 Economic Disruption

	General economic significant disruption						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Color	0.191***	0.169***	0.143***	0.131***	0.140***	0.137***	0.108***
	(0.0205)	(0.0219)	(0.0219)	(0.0220)	(0.0223)	(0.0220)	(0.0220)
RTW			0.178***				0.108***
			(0.0211)				(0.0224)
Acc Gvt Fund				0.217***			0.129***
				(0.0212)			(0.0252)
BK Relat					0.184***		0.0594**
					(0.0240)		(0.0278)
Fund Appl						0.191***	0.0765***
						(0.0209)	(0.0259)
Urban		0.0649***	0.0599***	0.0542**	0.0587***	0.0545**	0.0493**
		(0.0227)	(0.0222)	(0.0223)	(0.0226)	(0.0223)	(0.0220)
DSP		0.0333	0.0368	0.0336	0.0340	0.0270	0.0333
		(0.0290)	(0.0286)	(0.0281)	(0.0285)	(0.0283)	(0.0278)
Constant	0.365***	0.306***	0.205***	0.248***	0.280***	0.250***	0.180***
	(0.0145)	(0.0300)	(0.0313)	(0.0293)	(0.0295)	(0.0296)	(0.0305)
Observations	2,291	2,291	2,291	2,291	2,291	2,291	2,291
R-squared	0.037	0.040	0.069	0.084	0.064	0.075	0.105

NOTE: Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

are positive and significant, between 18 and 22 percent, indicating that even within communities with similar demographic compositions, communities that were facing these challenges were more likely to be severely disrupted. Most importantly, the coefficient on *Color* decreases between 15 and 22 percent when incorporating the challenges as explanatory variables of disruption. This is not surprising given that Figure 3 presents evidence that communities of Color were more likely to face serious hurdles. When all four challenges are considered, the coefficient on *Color* falls by approximately 35 percent. In other words, the observed differences in the severeness of the obstacles faced by the different demographic groups can explain 35 percent of the disruption gap.

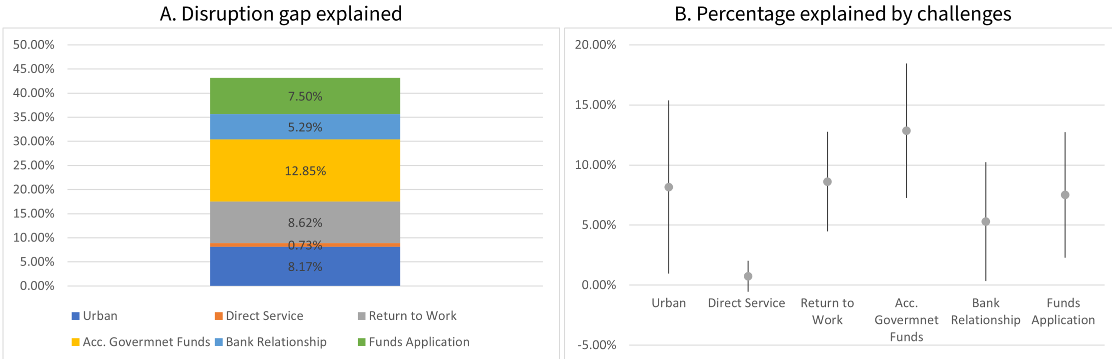
3.2.1 Relative Importance of Challenges

In this section I perform an Oaxaca decomposition to assess the relative importance of each challenge in explaining the disruption gap between White communities and communities of Color.

Of the 19-percentage-point difference in disruption between communities of Color and White communities (disruption gap of 19 percentage points), 8 percentage points can be explained by location, direct or indirect service, and all four challenges—this would account for 43 percent of the disruption gap. Figure 4 shows that approximately 13 percent of the disruption gap can be explained by unequal access to government relief plans and almost 9 percent by challenges returning to work. Lack of a bank relationship explains around 5 percent of the disruption gap, while barriers in applying for funds explains 7.5 percent. All four challenges can explain approximately 35 percent of the disruption gap, and location differences account for another 8 percent. While these numbers suggest that unequal access to government funding had a larger role in explaining the disruption gap, I cannot reject that this barrier was as relevant as the other ones (Figure 4B). In other words, I cannot claim that this obstacle was more important than the other challenges considered in explaining the disruption gap.

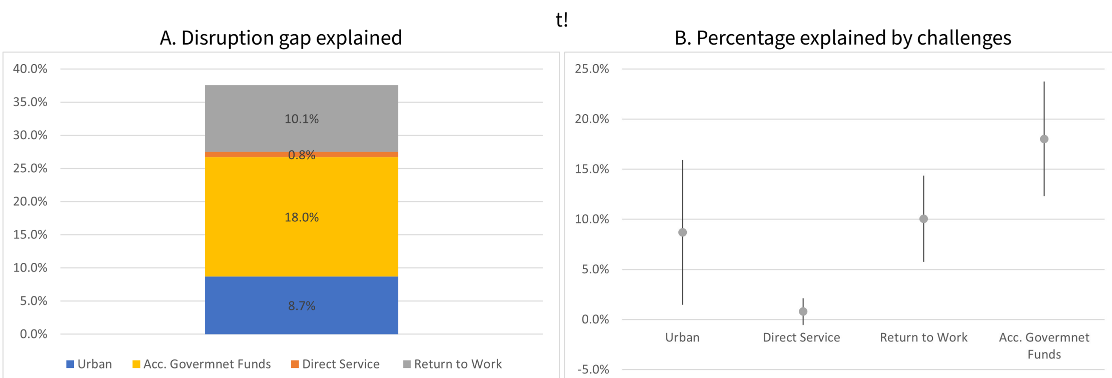
One could think that unequal access to government funding encompasses having difficulty applying for funds or lacking bank relationships to access funds since, in the end, communities either received funds late or not at all. Thus the correlation between these challenges might be reducing the overall importance of funding

Figure 4
Disruption Gap: Explanatory Power of Challenges



NOTE: The figure shows the percentage of the disruption gap explained by each variable of interest (panel A) and 95 percent confidence intervals for each variable of interest (panel B).

Figure 5
Returning to Work versus Unequal Access to Government Funding



NOTE: The figure shows the percentage of the disruption gap explained by each variable of interest (panel A) and 95 percent confidence intervals for each variable of interest (panel B).

relative to return-to-work challenges.[8] To address this concern, I test the robustness of my earlier findings by running equation (2) again but now having *Return to Work* and *Acc Govt Fund* as explanatory variables:

$$SigDisrup = \alpha + \theta * Color + \beta_1 * RTW + \beta_2 * Acc.Govt + \gamma * Urban + \delta * DSP + \upsilon. \tag{3}$$

Table 4 presents the results. Return-to-work barriers on their own explain approximately 15 percent of the gap (Column 2), while unequal access to government funding explains 22 percent of the gap on its own (Column 3). Together, they explain 30 percent of the gap (Column 5).

3.3 Returning to Work versus Funding

To test the relative weight of returning to work versus access to government funding, I perform an Oaxaca decomposition. I find that unequal access to government funding is as important as barriers to returning to work in shaping the disruption gap. While unequal access to government funding increases its impact on the gap, on average, so do barriers to returning to work (Figure 5). Their confidence intervals still overlap (Figure 5B), and thus unequal access to government funding remains as relevant as return-to-work barriers in shaping the disruption gap.

Next, I create a dummy variable, *All Funding*, that encompasses all disruptions related to funding (lack of bank relationship, challenges accessing government funding, and difficulty applying for funds). Mainly, this variable equals one if any of the above mentioned implied a significant barrier and equals zero otherwise. Table 4, Columns 4 and 6 show that this new variable is very similar to *Acc Govt Fund*. The coefficient on *Color* barely

8. Responses indicating significant challenge are positively correlated with a coefficient between approximately 0.4 and 0.5.

Table 4
Returning to Work versus Funding

	General economic significant disruption					
	(1)	(2)	(3)	(4)	(5)	(6)
Color	0.169***	0.143***	0.131***	0.131***	0.119***	0.120***
	(0.0219)	(0.0219)	(0.0220)	(0.0219)	(0.0219)	(0.0219)
RTW		0.178***			0.126***	0.120***
		(0.0211)			(0.0220)	(0.0223)
Acc Govt Fund			0.217***		0.180***	
			(0.0212)		(0.0223)	
All Funding				0.216***		0.178***
				(0.0207)		(0.0221)
Constant	0.306***	0.205***	0.248***	0.214***	0.187***	0.163***
	(0.0300)	(0.0313)	(0.0293)	(0.0300)	(0.0306)	(0.0309)
Observations	2,291	2,291	2,291	2,291	2,291	2,291
R-squared	0.040	0.069	0.084	0.085	0.097	0.096

NOTE: Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Obs: 2,291; Controls: DSP and Urban

differs from those in Columns 3 and 5. The change in the disruption gap is fairly small when I compare *All Funding* to *Acc Govt Fund*.

Overall, this preferred specification provides supportive evidence that unequal access to government funding as well as challenges returning to work were important features shaping the observed disruptions in communities. Together, they can explain almost 30 percent of the disruption gap, and both challenges were as important in explaining differences in disruption.

3.4 Segments of the Economy Disruption Gap and Challenges

The survey asked not only asked about general economic conditions but also about COVID-19 disruptions on communities along different segments of the economy—financial stability, small business, access to healthcare, services for children, housing stability, and basic consumption needs. All these segments of the economy are vital for vulnerable communities. As mentioned earlier, serious disruption along these areas varied between 30 and 60 percent. Nevertheless, across all segments, it was more likely for communities of Color to be significantly disrupted. Disruption gaps also varied across segments, with financial stability and services for children having the largest gaps: 22 and 16 percentage points, respectively.

I next examine the role that challenges played in shaping these observed gaps. Table 5 shows the disruption gap along the different sectors as well as the percentage of the gap explained by challenges. Challenges and location explain approximately 40 percent of the disruption differences in general economic disruption, financial stability, services for children, and basic consumption needs. More interestingly, they can be held accountable for most of the disruption gap in small business and housing stability, explaining 91 percent and 71 percent of the disruption gap, respectively. It is worth noting that the disruption gap in small business is small relative to the rest of the sectors analyzed.

Appendix Figure A.5 compares each one of the challenges to assess their relative importance in each segment. All four challenges were important and equally relevant in shaping the observed disruption gap for all six sectors except for services for children, where, reasonably, bank relationships seem to have no explanatory power on the disruption gap. In other words, the data suggest that challenges returning to work, lack of a bank relationship, unequal access to government funding, and barriers applying for funds all seem to be important in shaping the differences in the observed disruption across demographic groups. Together with the differences in community location, these challenges can explain more than 40 percent (up to 90 percent for certain sectors) of the observed disruption gap. These findings highlight the importance of the implementation and distribution

Table 5
Segments of the Economy and the Disruption Gap

	Disruption gap	Percentage of gap explained
General disruption	19.1%	43.2%
Financial stability	21.9%	43.3%
Small business	4.2%	90.8%
Access to healthcare	10.8%	57.8%
Services for children	16.4%	41.5%
Housing stability	10.9%	71.2%
Basic consumer needs	15.2%	42.8%

of COVID-19 government relief packages as well as the relevance of childcare, transportation, and healthcare in our economy.

3.5 Disruption Gap and Income

In this section I examine whether some of the racial disparities observed in the results stem from income differences in LMI communities and find that they do not. This is because income differences do not explain the variation in the disruption levels or the variation in the severity of the challenges faced. Moreover, there are income differences between White communities and communities of Color, even when comparing among LMI communities.

Matching communities' median household income to this data set offers some challenges as I do not have precise information on where communities were located and know only the location of the respondents' organization headquarters. A report by Faulk et al., 2021 shows that headquarter location is a strong predictor of the area served. That said, in some cases organizations serve in multiple locations as well as both regionally and nationwide. Using organizations' headquarter zip code as a proxy for the community's location will provide a noisy measure of income in the communities served, given that some organizations serve in several locations. Nevertheless, it is only way I can measure income in the community served given the available data.

Appendix Tables A.4 and A.5 show that incorporating communities' *Income*—logarithm of median household income—does not affect the coefficient or significance of *Color*. In other words, median household income does not alter the role that demographic composition plays in explaining general economic COVID-19 disruptions within LMI communities. This finding is robust when having a dummy variable for high income instead. *High Income* is a dummy variable that equals one when median household income is above the sample median and equals zero otherwise. In all cases, incorporating income does not alter the coefficient on *Color* or the relevance of the main challenges analyzed in explaining variation in disruption.

While this finding might be somewhat surprising, Appendix Table A.6 shows that variation in income cannot explain the differences in general economic disruptions, challenges returning to work, or accessing government funding. A possible explanation for this is that while there might be some variation in median income within my sample, the survey was already targeted at low-income communities, where job losses and financial instability have hit the majority of low-income neighborhoods in a similar manner.

Overall, these last two sections provide some robustness to the core of this article supporting the relevance of unequal access to government funding as well as the challenges of returning to work as important features shaping the observed disruptions in communities. The importance of these two indicators remains regardless of accounting for income or the segments of the economy analyzed, all important for the vitality of LMI communities.

4. CONCLUSION

The federal government responded to the COVID-19 pandemic with a safety net that went a long way in preventing widespread hardship (Cooney and Shaefer, 2021). The relief measures it provided reduced poverty, helped people access health coverage, and reduced hardships such as the inability to afford food or meet other basic needs. Funds and resources were substantial for LMI communities, with almost two-thirds of respondents indicating that stimulus checks, unemployment benefits, and rent relief were very critical during COVID-19 (Chalise and Gutkowski, 2021).

However, as this article finds, the ability to access funds has been unequal between communities of Color and White communities and can explain the observed differences in disruption levels across demographic groups. Possible reasons for these differentiated outcomes likely rely on historical inequities that cannot be easily resolved by distributing funds. Special attention and targeting might be necessary for funds to reach selected recipients, and further research on this targeting and implementation could help address these issues.

I also find that unequal access to government relief was as important as return-to-work challenges in shaping the observed demographic differences. Together, these challenges can explain approximately 30 percent of the disruption gap. While much research has been devoted to explaining how barriers to employment have had a destabilizing effect during COVID-19, this article highlights that the demographic variation in return-to-work challenges can explain part of the disruption gap between White communities and communities of Color. The findings highlight the importance of targeting, implementing, and distributing government relief packages as well as the relevance of childcare in our economy. Further work is needed to allow a more-inclusive labor market, such as one with accessible and affordable childcare.

REFERENCES

Autor, David, David Cho, Leland D. Crane, Mita Goldar, Byron Lutz, Joshua Montes, William B. Peterman, David Ratner, Daniel Villar, and Ahu Yildirmaz. 2022. The $800 billion paycheck protection program: where did the money go and why did it go there? *Journal of Economic Perspectives* 36, no. 2 (May): 55–80. https://doi.org/10.1257/jep.36.2.55. https://www.aeaweb.org/articles?id=10.1257/jep.36.2.55.

Brooks, Matthew M, J Tom Mueller, and Brian C Thiede. 2021. Rural-urban differences in the labor-force impacts of covid-19 in the united states. *Socius* 7:23780231211022094.

Budget, Center On, and US Policy Priorities. 2021. Tracking the covid-19 economy's effects on food, housing, and employment hardships. *Center on Budget and Policy Priorities. United States.[Web Archive] Retrieved from: https://www. cbpp. org/research/poverty-andinequality/tracking-the-covid-19-economyseffects-on-food-housing-and.*

Chalise, Nishesh, and Violeta A Gutkowski. 2021. Perspectives from main street: the impact of covid-19 on communities and the entities serving them. *Retrieved from Federal Reserve System Fed Communities,* https://fedcommunities.org/data/main-street-covid19-survey-2021.

Cooney, Patrick, and H Luke Shaefer. 2021. Material hardship and mental health following the covid-19 relief bill and american rescue plan act. *Ann Arbor, Michigan: Poverty Solutions, University of Michigan.*

Faulk, Lewis, Mirae Kim, Teresa Derrick-Mills, Elizabeth Boris, Laura Tomasko, Nora Hakizimana, Tianyu Chen, Minjung Kim, and Layla Nath. 2021. Nonprofit trends and impacts 2021. *The Urban Institute.*

Gangopadhyaya, Anuj, Michael Karpman, and Joshua Aarons. 2020. As the covid-19 recession extended into the summer of 2020, more than 3 million adults lost employer-sponsored health insurance coverage and 2 million became uninsured. *Washington, DC: Urban Institute.*

Gemelas, Jordan, Jenna Davison, et al. 2022. Inequities in employment by race, ethnicity, and sector during covid-19. *Journal of racial and ethnic health disparities* 9 (1): 350–355.

Granja, João, Christos Makridis, Constantine Yannelis, and Eric Zwick. 2022. Did the paycheck protection program hit the target? *Journal of financial economics* 145 (3): 725–761.

Holtzblatt, Janet, and Michael Karpman. 2020. Who did not get the economic impact payments by mid-to-late may, and why? *Washington, DC: The Urban Institute.*

Janger, Emma, Nicole Rubin, and Sejal Singh. 2020. Making unemployment insurance work for working people. *UCLA L. Rev. Discourse* 68:102.

Lee, Emma K, and Zachary Parolin. 2021. The care burden during covid-19: a national database of child care closures in the united states. *Socius* 7:23780231211032028.

Li, Lei, Philip E Strahan, and Song Zhang. 2020. Banks as lenders of first resort: evidence from the covid-19 crisis. *The Review of Corporate Finance Studies* 9 (3): 472–500.

Liu, Sifan, and Joseph Parilla. 2020. New data shows small businesses in communities of color had unequal access to federal covid-19 relief. *Brookings Institution Report.*

Maxwell, Connor. 2020. The coronavirus crisis is worsening racial inequality. *Center for American Progress website: americanprogress. org/issues/race/news/2020/06/10/486095/coronavirus-crisis-worsening-racial-inequality. Published June* 10.

OECD. 2022. The unequal impact of covid-19: a spotlight on frontline workers, migrants and racial/ethnic minorities, https://doi.org/https://doi.org/https://doi.org/10.1787/f36e931e-en. https://www.oecd-ilibrary.org/content/paper/f36e931e-en.

Parker, Kim, Juliana Menasce Horowitz, and Rachel Minkin. 2021. Americans are less likely than before covid-19 to want to live in cities, more likely to prefer suburbs.

APPENDIX 1.

Table A.1
COVID-19's Level of Disruption Segments of the Economy

	No disruption (%)	Some disruption (%)	Significant disruption (%)	Unsure (%)	Observations
General disruption	1.4	51.2	46.4	1.0	2,391
Financial stability	1.7	41.3	54.3	2.8	2,291
Small business	1.1	34.7	60.4	3.7	1598
Access to healthcare	4.6	43.0	40.8	11.6	2,288
Services for children	1.8	29.1	59.1	10.0	2,287
Housing stability	3.4	31.0	55.3	10.3	2,289
Basic consumer needs	5.5	51.9	35.9	6.7	2,288

Table A.2
Disruption by Color, Type of Service, and Location

	General economic disruption				
	No disruption (%)	Some disruption (%)	Significant disruption (%)	Unsure (%)	Observations (% of total)
Color: No	2.5	60.5	36.3	0.8	48
Color: Yes	0.4	42.5	55.8	1.2	52
No DSP	2.7	55.9	40.8	0.7	14
DSP	1.2	50.5	47.3	1.1	86
Rural	2.1	59.5	37.9	0.5	35
Urban	1.1	46.8	50.9	1.3	65
Total obs.	31	1,115	1,010	22	

NOTE: This table shows the distribution of responses to general economic disruption disaggregated by communities' demographics, type of service provided, and location.

Table A.3
Distribution of Challenges among Respondents

	Challenges			
	RTW (%)	Acc Gvt Fund (%)	BK Relat (%)	Fund Appl (%)
Not a challenge	2.2	7.5	14.3	6.4
Challenge	27.8	41.9	29.6	36.5
Significant challenge	65.9	40.1	27.3	46.4
Unsure	4.1	10.6	28.8	10.6
Observations	2,222	2,233	2,019	2,195

Table A.4
Robustness: Controlling for Income

	General economic significant disruption					
	(1)	(2)	(3)	(4)	(5)	(6)
Color	0.195***	0.195***	0.194***	0.170***	0.170***	0.169***
	(0.0210)	(0.0210)	(0.0210)	(0.0226)	(0.0226)	(0.0226)
Income		-0.000126			-0.00267	
		(0.0258)			(0.0260)	
High Income			-0.0112			-0.0143
			(0.0210)			(0.0210)
Urban				0.0691***	0.0693***	0.0704***
				(0.0233)	(0.0234)	(0.0234)
DSP				0.0472	0.0469	0.0458
				(0.0301)	(0.0302)	(0.0302)
Constant	0.363***	0.364	0.369***	0.290***	0.319	0.298***
	(0.0148)	(0.285)	(0.0186)	(0.0310)	(0.289)	(0.0332)
Observations	2,178	2,178	2,178	2,178	2,178	2,178
R-squared	0.038	0.038	0.038	0.043	0.043	0.043

NOTE: Robust standard errors are in parentheses.*** p<0.01, ** p<0.05, * p<0.1

Table A.5
Robustness: Controlling for Income (2)

	General economic significant disruption		
	(1)	(2)	(3)
Color	0.111***	0.111***	0.110***
	(0.0227)	(0.0227)	(0.0227)
Income		0.00203	
		(0.0248)	
High Income			-0.00933
			(0.0204)
Urban	0.0533**	0.0532**	0.0542**
	(0.0227)	(0.0227)	(0.0227)
DSP	0.0480*	0.0482*	0.0471
	(0.0289)	(0.0290)	(0.0290)
RTW	0.110***	0.110***	0.110***
	(0.0230)	(0.0230)	(0.0230)
Acc Govt Fund	0.124***	0.124***	0.124***
	(0.0258)	(0.0258)	(0.0258)
BK Relat	0.0527*	0.0528*	0.0521*
	(0.0285)	(0.0285)	(0.0285)
Fund Appl	0.0786***	0.0785***	0.0786***
	(0.0266)	(0.0266)	(0.0265)
Constant	0.164***	0.142	0.169***
	(0.0313)	(0.275)	(0.0334)
Observations	2,178	2,178	2,178
R-squared	0.105	0.105	0.105

NOTE: Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A.6
Disruption, Challenges, and Income Differences

	Color		General sig. disrup.		RTW sig. chall.		Acc Govt sig. chall.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Income	-0.0540**		-0.0106		0.00110		-0.0159	
	(0.0260)		(0.0264)		(0.0248)		(0.0260)	
High Income		-0.0367*		-0.0184		-0.00735		-0.00918
		(0.0214)		(0.0214)		(0.0206)		(0.0209)
Constant	1.110***	0.535***	0.580**	0.473***	0.627**	0.643***	0.565**	0.395***
	(0.285)	(0.0151)	(0.290)	(0.0151)	(0.273)	(0.0145)	(0.286)	(0.0148)
Observations	2,178	2,178	2,178	2,178	2,178	2,178	2,178	2,178
R-squared	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000

NOTE: Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure A.1
Demographics and Communities Served

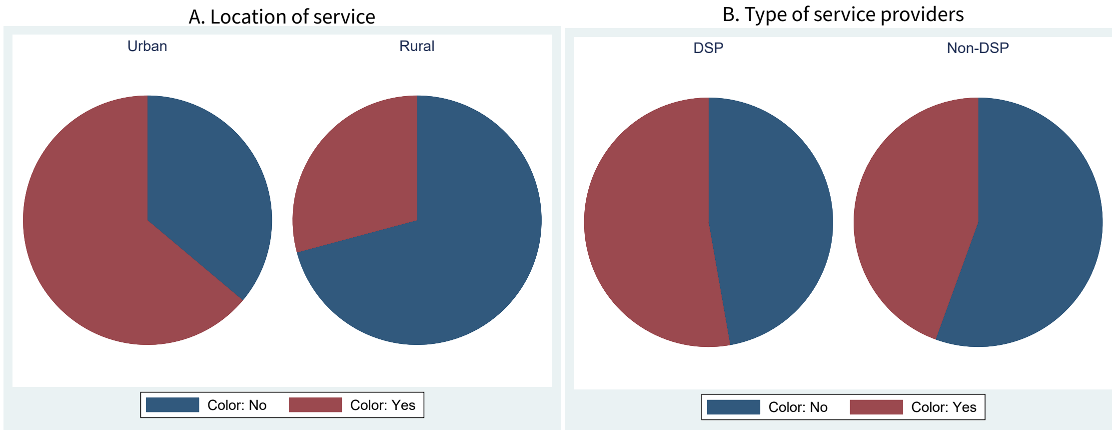


Figure A.2
Disruption by Type of Service Offered and Location

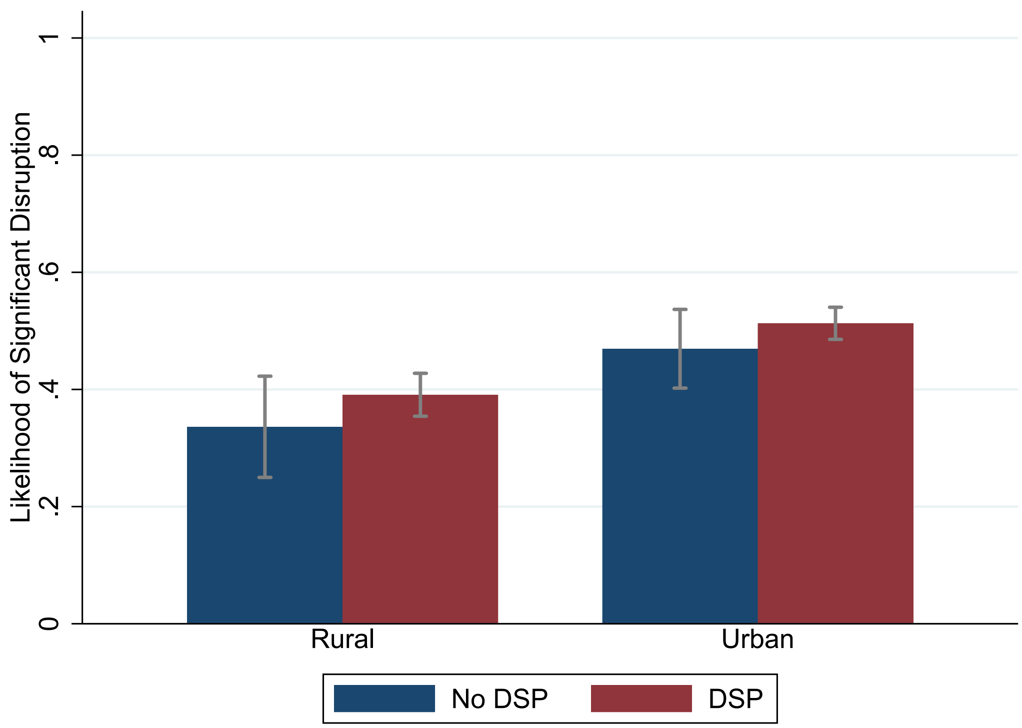


Figure A.3
Disruption in Segments of the Economy

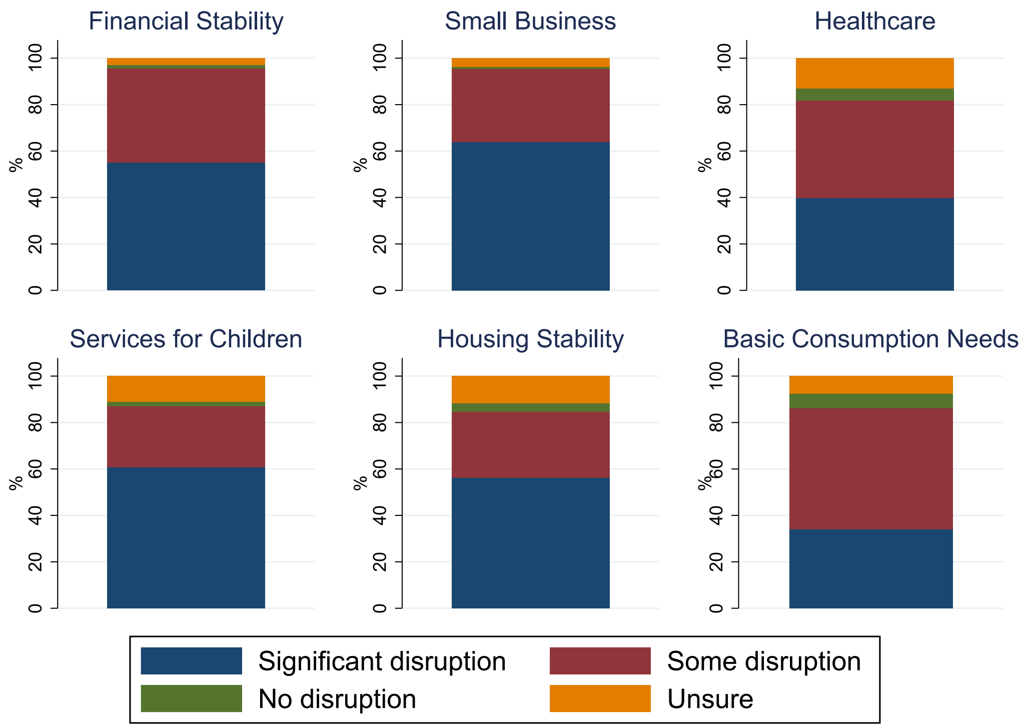


Figure A.4
Challenges Faced by Demographic Group

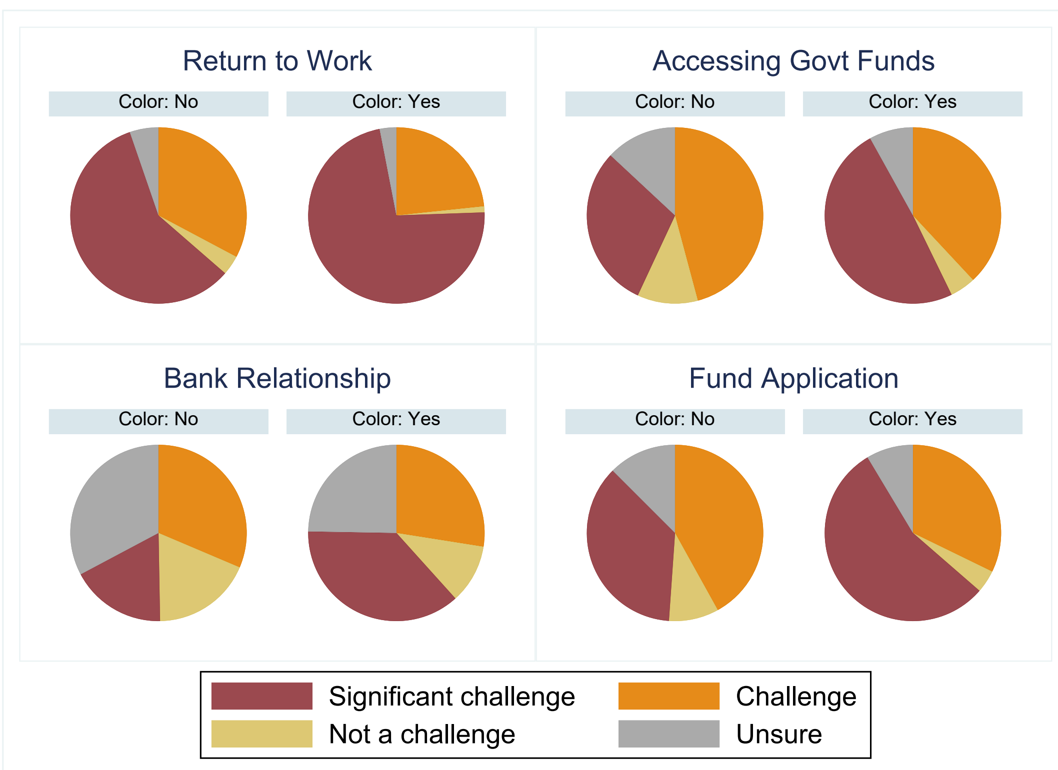
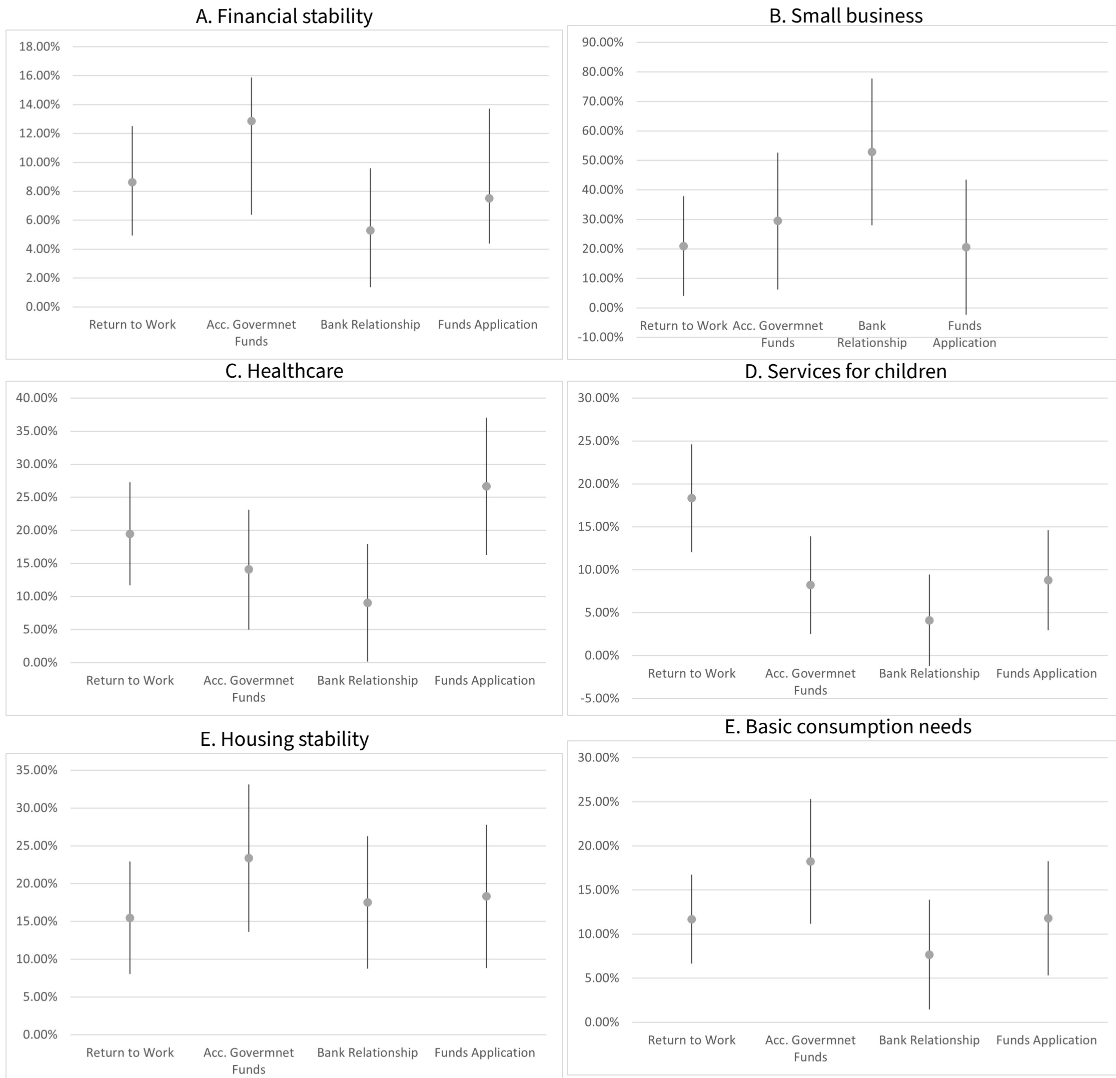


Figure A.5

Confidence Intervals of Challenges by Segments of the Economy



NOTE: The figures show 95 percent confidence intervals from the Oaxaca decomposition for each of the challenges analyzed. There is one figure for the decomposition in each segment of the economy.

An Introduction to Zero-Knowledge Proofs in Blockchains and Economics

Aleksander Berentsen, Jeremias Lenzi, and Remo Nyffenegger

Abstract

With a zero-knowledge proof (ZKP), a party can prove that a statement is true without revealing any information except for whether it is indeed true or not. The obvious benefit is privacy since the prover does not need to reveal any additional information, and the second benefit is that it can significantly reduce the cost of verifying the correctness of a statement. ZKPs are increasingly adopted in blockchain applications, where privacy and efficiency still have a lot of room for improvement. While it is expected that ZKP technology will also become ubiquitous in many other areas, the term remains cryptic to many people without a computer science background. In this review article, we shed light on what ZKPs are and how they improve privacy and efficiency and describe applications for blockchains and other use cases.

JEL codes: E3, E52, E62, F10, O10

Federal Reserve Bank of St. Louis *Review*, Fourth Quarter 2023, *105*(4), pp. 280-94.
https://doi.org/10.20955/r.105.280-94

1. INTRODUCTION

Efficiency in economics means that scarce resources should not be wasted. This statement is true for manufacturing as well as for computations since they use real resources such as hardware, electricity, or human capital. There are various applications in which a computation's correctness must be verified by many other parties. For example, in blockchains, decentralization requires that many network participants recompute the correctness of each block that is appended to the chain.[1] Obviously, reexecuting the same computations is inefficient because it involves using computational resources repeatedly.

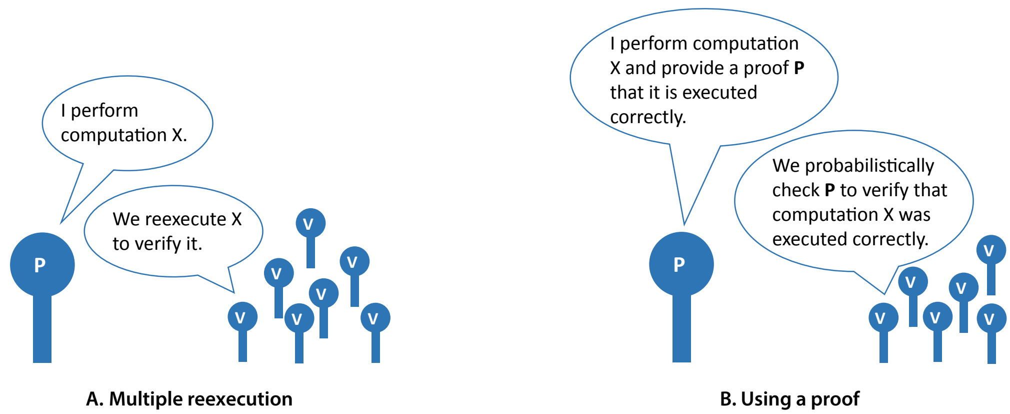
By using a zero-knowledge proof (ZKP), a party can prove to other parties that a computation was executed correctly. There is no need to replicate the computation—only the proof needs to be verified. Ideally, verifying a ZKP needs significantly less resources than reexecuting the computation. This benefit is illustrated in Figure 1; note that the efficiency gains of ZKPs increase linearly in the number of validators.

The second and more obvious benefit of ZKP technology is privacy. By using a ZKP, one can prove the correctness of a computation without revealing any additional information except for whether it is indeed correct or not. For example, a blockchain user can prove that he is indeed allowed to make a payment without revealing his identity to the network. Existing applications are the privacy-protecting cryptocurrency Zcash and the Tornado cash protocol on Ethereum (see Nadler and Schaer, 2023). The privacy and confidentiality

1. To learn more about blockchain technology, see Schaer and Berentsen, 2020.

Aleksander Berentsen is a professor of economic theory at the University of Basel and a research fellow at the Federal Reserve Bank of St. Louis. Jeremias Lenzi is a PhD student at the University of Basel. Remo Nyffenegger is a PhD student and research assistant at the Center for Innovative Finance at the University of Basel.

Figure 1
Reexecution vs. Using a Proof



of data is also important outside of blockchains. Two examples are a person who wants to prove that she voted without revealing her vote, or a company that wants to prove its solvency without revealing its balance sheet.

The theoretical concept of ZKPs was introduced in the late 80s by Goldwasser, Micali, and Rackoff, 1989. Conceptually, there are many different use cases, but none of them have become economically important. This has changed with the advent of blockchain technology, where ZKP technology has been integrated in some applications. ZKP research is rapidly expanding as demonstrated by the increasing number of articles about the technology (see Burger et al., 2022 for an overview).

Most research on ZKPs targets an audience with a computer science or mathematics background, and there is a lack of a comprehensive but intuitive introduction into the topic. This review article fills this gap by providing an accessible but extensive introduction into zero knowledge proofs and their applications. Furthermore, in Berentsen, Lenzi, and Nyffenegger, 2022 we provide a comprehensive example of a ZKP that includes more advanced math and comes with an accompanying Python script.

The rest of the article is organized as follows. In Section 2 we discuss two very simple and intuitive examples of a ZKP, while in Section 3 we dive more into the details. In Sections 4 and 5 we discuss the application of ZKPs for blockchains and in general, and in Section 6 we conclude.

2. INTUITIVE EXAMPLES OF A ZKP

From a technological perspective, we can differentiate two types of proofs, interactive and non-interactive ones. An interactive proof requires some bilateral interaction between the prover and verifier; i.e., the verifier sends personalized random queries/challenges to the prover, who then responds to these queries. This process is repeated over several rounds until the verifier is convinced of the correctness of the proof with a sufficiently high probability. Each additional verifier who also wants to check the proof has, again, a bilateral interaction with the prover and sends personalized challenges over several rounds.

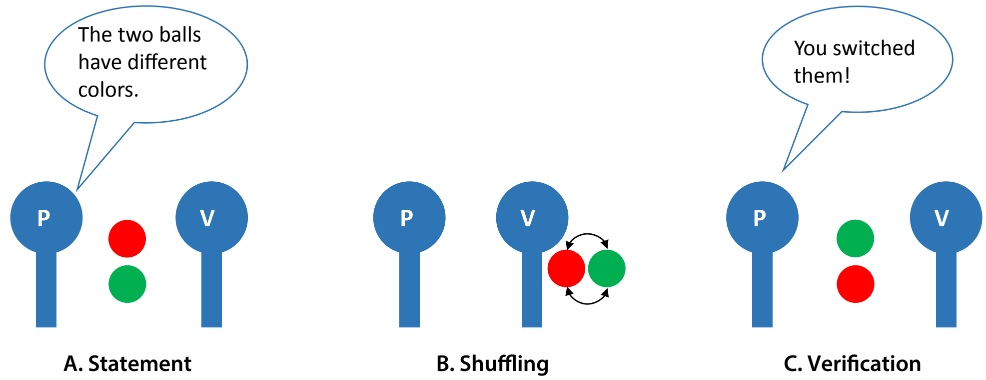
Thus, an interactive proof can be inefficient or infeasible if many verifiers are involved.[2] In a non-interactive proof, no interaction is required and the prover provides only one proving object (e.g., a string) that anybody can verify at any time. This is useful in, for example, blockchain applications. Below, we provide one simple example for both types of proofs: first for an interactive proof and then for a non-interactive one.

2.1 Two Balls and the Colorblind Friend

To get a first intuition of a ZKP, we will discuss a trivial example by Chalkias and Hearn, 2019, which is called "two balls and the colorblind friend." Imagine two friends, Peggy and Victor. Peggy is an expert in ZKPs and wants to teach the concept to Victor. She knows that Victor is colorblind and hands him two identical balls except one is green and the other is red. They agree that the balls are exactly identical apart from their color, which Victor cannot evaluate because of his colorblindness. Peggy (the prover) claims that she can prove to Victor (the verifier) that the balls have indeed different colors without revealing which is which (i.e., the zero-knowledge part). This example is illustrated in Figure 2A.

2. There are two main reasons: generating the proof repeatedly increases the overall computing power used, and a prover must be available/online every time a verifier wants to verify the proof.

Figure 2
The Two Balls and the Colorblind Friend



The proof in this example works like a small game. Victor takes the two balls, puts them behind his back, shuffles them (Figure 2B), and then again shows them to Peggy (Figure 2C). He knows whether he has switched the balls or not since he shuffled them consciously. Furthermore, Peggy knows as well because she can differentiate between the colors. She then tells him that he switched, which is an indication that the balls have to be differently colored.

The proof, however, does not end here. Victor might think that Peggy was just lucky and correctly guessed by chance. The probability of doing so is indeed 50 percent, and therefore he repeats the experiment. An excellent mathematician, he knows that the probability of having two balls with different colors is $P(\text{balls have two colors}) = 1 - 0.5^n$, where n is the number of times they repeat the game. Already after $n = 10$ iterations, the probability that the balls are indeed differently colored is $P = 99.9$ percent. Thus, Peggy can prove her claim that the balls have different colors without revealing which ball has which color. If she intends to convince other people of the proof, she would need to repeat the same steps for each verifier.

2.2 Where's Waldo?

We all know the *Where's Waldo* books with the illustrations of dozens of people in a chaotic environment, where the goal is to find Waldo in the red and white striped jumper. Peter is a very talented Where's Waldo player. After the release of a new book, he wants to prove to the *Where's Waldo*-community that he has found Waldo in an especially difficult illustration. Obviously, he does not want to reveal where exactly Waldo is. We slightly extend the concept of Naor, Naor, and Reingold, 1999 to show how Peter can prove the statement "I have found Waldo" using a form of a non-interactive ZKP proof.

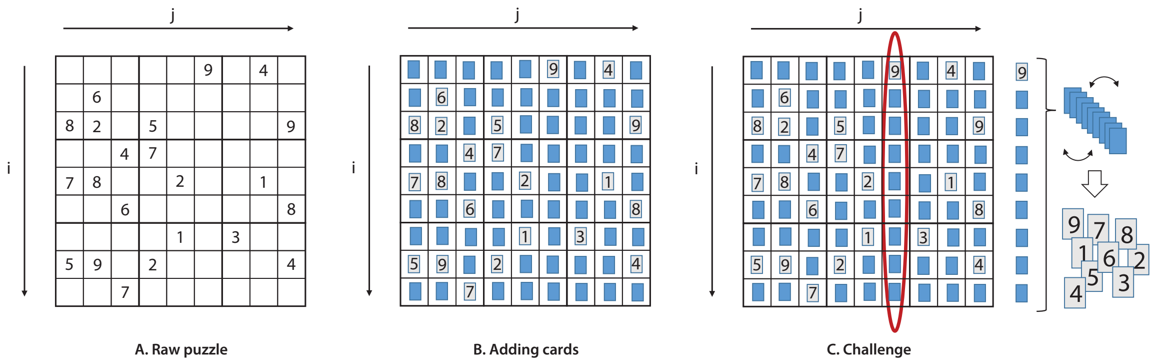
Before the actual proof, the community and Peter agree on the rules.[3] They define an empty room with nothing in it except for the illustration, a photocopy of it, and a pair of scissors. The proof begins with Peter being searched before entering the room to make sure that he does not take anything with him. This arrangement is called an initial set-up or ceremony. All community members attending the ceremony can verify themselves that Peter cannot cheat because the room indeed only contains the illustration and a pair of scissors and Peter was searched properly. He could only cheat if all community members attending the set-up collude. Hence, a community member who did not attend the ceremony can also be convinced of the proof if at least one attendant in the ceremony is honest.[4]

Next, Peter is alone in the room and cuts out Waldo from the photocopy and puts him next to the illustration. He then leaves the room, taking the remaining parts of the photocopy with him. The cropped image of Waldo is the only object that Peter provides as a proof. Now each member of the community can verify the proof by entering the room and checking whether the statement that Peter found Waldo is true. The proof is non-interactive since the verifiers do not need to interact with Peter but can just verify the cropped image.

3. In more complex examples, the prover and the verifier must agree up front on certain conditions that need to hold for the proof to be true.

4. Also, some types of real-world, non-interactive ZKP implementations require these ceremonies. As in our example, as long as at least one party in the ceremony acts benevolently, the subsequent proofs can be trusted. We describe this in more detail in Section 3.3.

Figure 3
Sudoku



3. ADVANCED EXAMPLE AND SOME THEORY

After providing some intuition on ZKPs, we now dive in deeper. A ZKP must fulfill three properties:

- Zero knowledge: A malicious[5] verifier cannot extract any information except for the statement being true or not.
- Completeness: A prover can convince the verifier (who follows the protocol) of a true statement.
- Soundness: A malicious prover cannot convince the verifier of a false statement with a sufficiently high probability.

The examples above satisfy all properties. The two-ball example is zero knowledge because the game structure (the protocol) does not reveal to Victor which ball has which color and he cannot derive it in a different way because of his colorblindness. Completeness is achieved because Peggy can convince Victor that the balls are indeed differently colored. With respect to soundness, a false statement would be that Peggy claims that the balls have different colors even though they are actually identical in that perspective as well.

By playing the game with identically colored balls, Peggy has in each round a 50 percent chance of guessing correctly whether Victor switched the two balls or not. After 10 rounds, the probability of guessing correctly in each round is only 0.1 percent. Thus, Peggy cannot convince Victor of the false statement with a sufficiently high probability, and also soundness is satisfied. The Where's Waldo example is zero knowledge because the verifiers cannot extract any information about where Waldo is, it is complete because Peter can show where Waldo is, and it is sound because showing another character than Waldo would not convince the verifiers.

3.1 Sudoku

We now introduce a more advanced interactive proof of a Sudoku puzzle that follows Gradwohl et al., 2009. The main purpose is to gain more intuition without showing very complicated math. But compared to the examples before, we incorporate some additional features that a ZKP normally entails.

Remember that in a Sudoku game the player is given a 9×9 grid in which some cells are filled with a digit. The objective is to fill the grid such that each row, each column, and each of the nine 3×3 boxes contain all digits from 1 to 9. We denote the row as $i \in \{1, 2, \ldots, 9\}$, the column as $j \in \{1, 2, \ldots, 9\}$, and the respective digit in cell $[i, j]$ as $k \in \{1, 2, \ldots, 9\}$. Furthermore, we denote the boxes as $b \in \{1, 2, \ldots, 9\}$, counting from left to right and top to bottom.

Figure 3A illustrates an example of a Sudoku puzzle. Using the notation above, we can describe the puzzle by a set of vectors of the type (i, j, k). An example of a vector would be $(3, 4, 5)$, which means that in row 3 and column 4, the entry is 5. The puzzle in Figure 3A can thus be represented by a set of 22 vectors, one for each number in the grid.

Peggy is in a Sudoku competition with her friend Victor, and whoever solves the puzzle first wins. She is the first to find a solution and wants to prove to Victor that she solved it without revealing the solution; i.e., she wants to prove the statement "I have a solution to this specific Sudoku game," to which the answer is either true or false. To do so, she proposes to apply a form of a ZKP that only requires the puzzle to be in a big format and some decks of playing cards.

5. Malicious means that a party tries to cheat or trick the protocol.

The first step is to represent a filled puzzle in a 9-by-9-by-9 cube, denoted as x that contains binary values in each cell:

$$x(i,j,k) = \begin{cases} 1, & \text{if } k \text{ is the entry in cell } [i,j] \\ 0, & \textit{otherwise}. \end{cases}$$

Next, Peggy and Victor agree on some constraints that need to hold if the solution is correct:[6]

$$\forall j,k : \sum_{i=1}^{9} x(i,j,k) - 1 = 0, \tag{1}$$

$$\forall i,k : \sum_{j=1}^{9} x(i,j,k) - 1 = 0, \tag{2}$$

$$\forall U,V \in \{0,3,6\} : \sum_{i=1}^{3}\sum_{j=1}^{3} x(i+U, j+V, k) - 1 = 0. \tag{3}$$

Equation (1) means that column j contains only one appearance of the digit k, equation (2) represents the fact that in row i there must be only one value of k, and equation (3) states that in each of the nine 3×3 boxes, the digit k appears only once.[7] Note that if constraints (1), (2), and (3) hold, the solution to the puzzle is correct, and by verifying them, Victor does not learn in which cell is which value. This is important for the zero-knowledge property.

Peggy now takes the playing cards with the digits 1 to 9. On the cells with the initial numbers, she puts the corresponding playing cards on top with the digits facing upward.[8] The other cards that correspond to her solution are put facedown onto the puzzle (Figure 3B).

Now the interactive part of the proof begins. Victor randomly chooses either a row, column, or box—here, he selects column $j = 6$. Peggy collects all nine cards of this column, shuffles them, and hands them to Victor. Victor can then look at the cards and check whether constraint (1) holds, i.e., whether each digit from 1 to 9 appears only once. This part represents one round of interaction in the protocol and is illustrated in Figure 3C. Victor makes a query—i.e., randomly chooses a row, column, or box—and Peggy's response is handing him the respective cards, which he can then verify.

Obviously, querying only once is not enough because Peggy might have set up the digits correctly in this column without having a proper overall solution and was just lucky that Victor chose this column by chance. Thus, Victor repeats the procedure by choosing additional rows, columns, and boxes. In each round, Peggy takes the corresponding cards, shuffles them, and hands them to Victor, who then verifies that either equation (1), (2), or (3) holds.

With each querying round, the probability that Peggy has indeed found a proper solution increases and the soundness error decreases. Victor can make as many queries as he wants until he is convinced that the probability of Peggy's statement being false is sufficiently small. He could theoretically also query for all 27 possible elements such that the probability of being cheated becomes 0 percent. However, this takes more time and makes the proof bigger and inefficient. Thus, he can define which soundness error is acceptable for him and then choose the corresponding number of queries. Hence, there is a trade-off between succinctness[9] (proof size) and soundness (security). Depending on the ZKP algorithm, this trade-off can also be prevalent in more complex proofs.

The proof described above fulfills the properties of zero knowledge, completeness, and soundness. It is zero knowledge because it is not possible to derive which number is in which cell because Peggy shuffles the cards, and it is complete because she can convince Victor that she found a correct solution. It is also sound because if she cheated by setting up the numbers in a wrong way, he will catch her with a sufficiently high probability.

This example brings us a bit closer to a more realistic ZKP. We introduced constraints and showed how the prover and verifier interact, how the verifier makes random queries, and how he checks whether the constraints are true. However, we are still far from a proper ZKP in which much more math is involved; to learn more, see Berentsen, Lenzi, and Nyffenegger, 2022.

---

6. In this example it is easy to come up with mathematical constraints that relate to the statement. In more involved proofs, more complex constraints first need to be transformed to a form that a ZKP can deal with. See, e.g., Buterin, 2016 for an intuitive explanation.
7. The terms U and V allow to "jump" to the next box.
8. Victor will challenge for several rows, columns, or boxes, and hence Peggy needs to put several cards on top of each other. For the sake of simplicity, we here assume that she does not cheat by putting distinct cards on top of each other.
9. Succinctness means that the proof can be efficiently verified.

3.2 Making Sudoku Non-Interactive

So far, we have seen two interactive (two balls and Sudoku) and one non-interactive (Waldo) proofs. We now want to show how it is possible to transform an interactive proof into a non-interactive one, using the Sudoku example. Showing this is a bit complicated, but by doing so, we learn new concepts that are used in real non-interactive proofs. Specifically, we introduce cryptographic hash functions, commitments, and the Fiat-Shamir heuristic, which are all used in ZKPs. To keep this section as simple as possible, we will abstract partly from the zero-knowledge property. Hence, the proof is closer to a validity proof, which is more about proofing the correctness of the statement efficiently rather than the zero-knowledge part.

First, we need to introduce cryptographic hash functions and modular arithmetic. A cryptographic hash function is a deterministic function that is practically infeasible to invert. It takes an input and creates an output of fixed length. Below we illustrate a few examples using the CRC-16 hash function,[10] which is denoted by H and returns an output in hexadecimal[11] form:

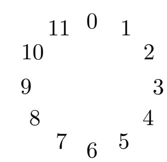
$$H(\text{zero-knowledge proof}) = ec65, \tag{4}$$

$$H(\text{blockchain}) = c964, \tag{5}$$

$$H(375869241) = fbd1, \tag{6}$$

$$H(348972651) = 4a33. \tag{7}$$

Modular arithmetic can be best illustrated by using a clock, and in this case the modulo operator is 12. Whenever an operation yields a result that is larger than or equal to 12, we follow the clock clockwise, starting with 0 again. The following are some examples:



$$11 + 2 \mod 12 = 1, \quad 29 \mod 12 = 5, \quad 6 \cdot 7 \mod 12 = 6$$

We saw in the interactive proof of the Sudoku puzzle that random queries by Victor are crucial. If they were not random and Peggy knew that the queries involved, for example, column $j = 6$, row $i = 9$, and box $b = 1$, she could have ordered the cards correctly only for these queries and cheated. To handle this in a non-interactive proof, we introduce commitments and use cryptographic hash functions to get pseudorandom queries. Furthermore, we stated above that in a non-interactive proof, Peggy provides only one proving object that everybody can use to verify the proof. In this example it is a string of characters denoted as $\mathcal{P}$, and we will show exactly how Peggy creates $\mathcal{P}$ and how a verifier can verify it. Figure 4 depicts the solution to the Sudoku puzzle.

The first part of the protocol requires Peggy to commit to the solution. She does so by applying a cryptographic hash function to all rows, columns, and boxes and including a compressed form of the hash outputs in the proof string. For example, she reads the numbers of the first row $i = 1$, i.e., 375869241, and applies the cryptographic hash function $H()$ to them. The result of this operation is $fbd1$ as shown in equation (6). The output of applying $H()$ to the numbers in column $j = 1$ is $4a33$ (see equation (7)).

Peggy then includes a compressed form of these hash outputs in the proof string whereby she commits to the solution. She cannot include the digits directly because she would reveal the solution. However, due to the deterministic nature of the cryptographic hash function, she can create a link from the solution to the hash output. Furthermore, by seeing only the hash output, another party cannot derive the solution because of the irreversibility property of the cryptographic hash function. The commitment ensures that Peggy cannot maliciously modify the solution in the subsequent steps of the proof; this will become clearer later.

Peggy could add all 27 hashes for all rows, columns, and boxes to the proof string $\mathcal{P}$. However, this would unnecessarily blow up the proof's size. Thus, she builds a Merkle tree as illustrated in Figure 5. In a Merkle tree, the elements on the bottom consist of the hashes of the rows, columns, and boxes. To go upward, the two elements just next to each other are hashed together in pairs until only one element is left, which is called the Merkle root. We again use the hash function CRC-16.[12] Peggy adds the root to the proof string, i.e., $\mathcal{P} = eb47$, and thereby commits to the solution. The root is the compressed form of the hash outputs mentioned in the paragraph above.

Now, remember that we need to introduce some pseudorandomness to replace the random queries of Victor in the interactive proof; i.e., we need to define how Peggy selects the first row, column, or box. The protocol solves this as follows:

10. https://emn178.github.io/online-tools/crc16.html
11. Hexadecimal is a numeral system that includes the symbols 0–9 and A–F. You can easily convert a hexadecimal number to a decimal or binary one.
12. https://emn178.github.io/online-tools/crc16.html

Figure 4
Sudoku Puzzle Solution

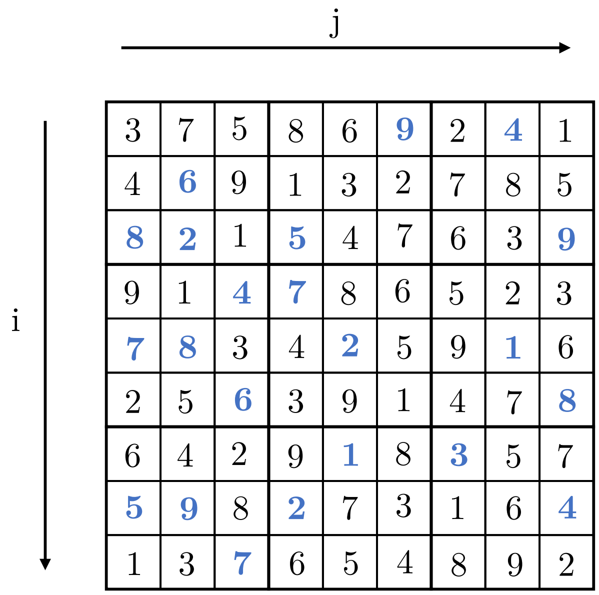
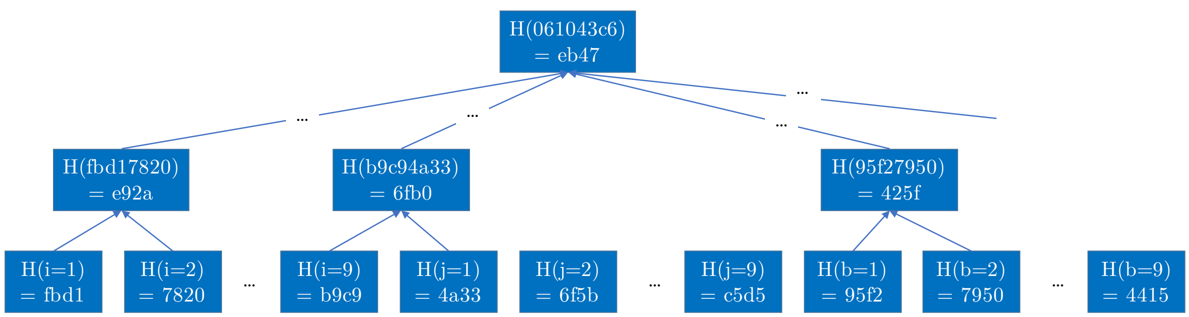


Figure 5
Merkle Tree



1. Peggy rewrites the Merkle root as a decimal number; i.e., $eb47 = 60231$.
2. She uses a mapping to transform the Merkle root in decimal form into $\mathbb{Z}_{27}$, where $\mathbb{Z}_{27} \in \{0, 1, \ldots, 26\}$. An easy way to do that is by applying modular arithmetic; i.e., $60231 \bmod 27 = 21$.
3. If the resulting number is between 0 and 8, Peggy matches it to rows $i \in \{1, 2, \ldots, 9\}$. If it is between 9 and 17, she matches it to columns $j \in \{1, 2, \ldots, 9\}$, and if it is between 18 and 26, she matches it to boxes $b \in \{1, 2, \ldots, 9\}$. Since the result before was 21, she selects box $b = 4$ for the first querying round.

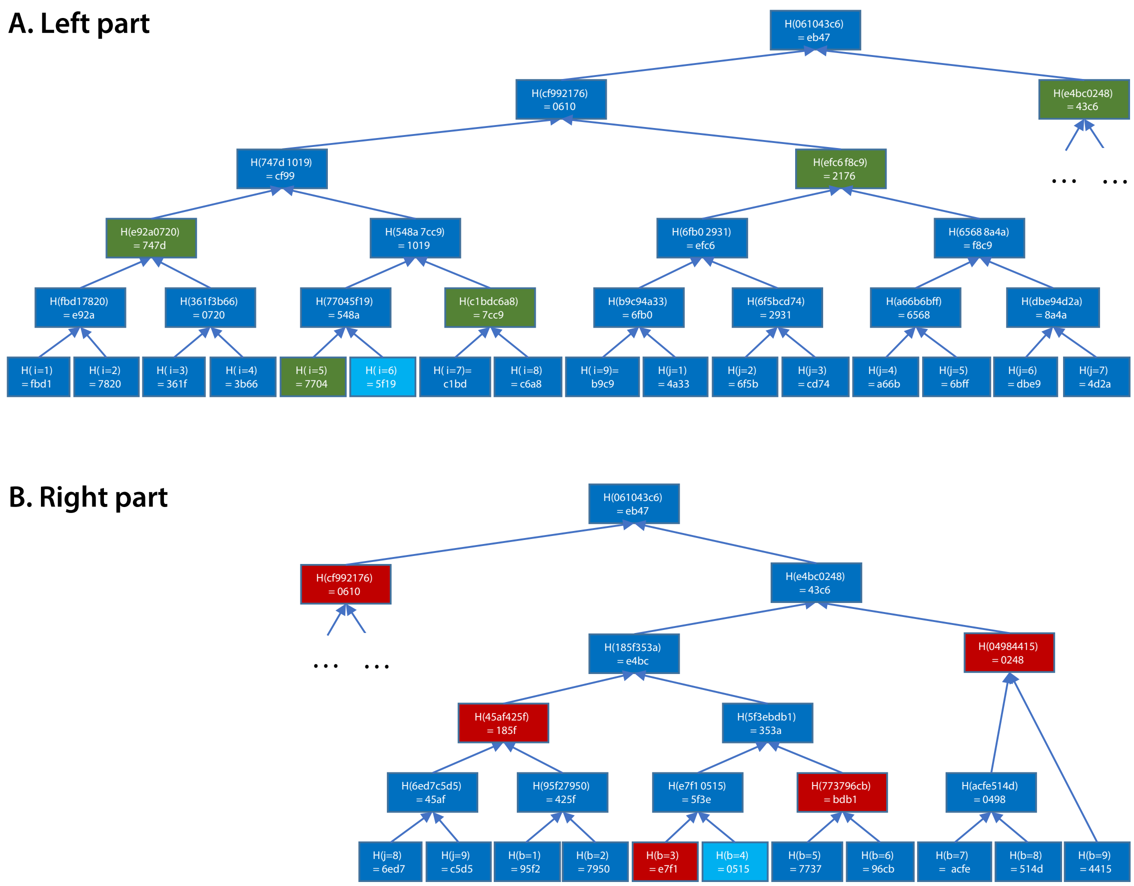
Replacing a random query by using a cryptographic hash function's pseudorandom output is called the Fiat-Shamir heuristic. This heuristic is also used in more complex proofs to make them non-interactive.

Peggy writes down the result of box $b = 4$, which is 914783256, and adds it to the proof string that already contains the Merkle root *eb*47; i.e., $\mathcal{P} = eb47\,\mathbf{914783256}$. Now the commitment becomes important. Peggy could have written down any random combination of the digits 1 to 9 that does not correspond to the correct solution. However, she can prove correctness by using the Merkle tree as illustrated in Figure 6.

When building the Merkle tree, she uses the solution in box $b = 4$ to get a leaf at the bottom with hash 0515. If she chooses any different sequence of digits, a different hash would result in the bottom leave. Consequently, the hashes in the upper leaves would change as well, and a different Merkle root would result. But since she uses the Merkle root to select the box in the first place, she would end up with another row, column, or box choice in the first round, making it quite hard for her to cheat.

A verifier can check whether Peggy provides the correct sequence of digits that corresponds to the solution committed in the Merkle tree as follows. The verifier first hashes the substring in $\mathcal{P}$ to get the bottom leaf; i.e., $H(914783256) = 0515$, represented by the light blue box in Figure 6B. He then uses the result 0515 and

Figure 6
Sudoku



hashes it pairwise with all the hashes of the red boxes in Figure 6b to reconstruct this path of the tree. Finally, he checks whether the resulting Merkle root equals the one in $\mathcal{P}$. If they are equal, he is convinced that Peggy did not choose any wrong digit combination.

For the verifier to be able to perform these steps, he needs to know the hashes in the red boxes. Thus Peggy must provide them in the proof string in the proper order; i.e., $\mathcal{P}$ = *eb*47 914783256 **e7f1 bdb1 185f 0248 0610**. This concludes the first querying round. Peggy used the initial values of the Sudoku puzzle to pseudorandomly determine which column, row, or box she should choose. She added the sequence of digits in this row to the proof string and also attached certain hashes of the Merkle tree to convince a verifier that she did not just write down any random digits into the proof string.

For all subsequent rounds, we define the protocol as such that the pseudorandomness comes from the solution in the last round. Thus, the second querying round would work as follows. First, Peggy calculates the solution's hash in the last round; i.e., $H(914783256) = 0515$. She then rewrites it as a decimal number, i.e., $0515 = 1301$, and applies the mapping; i.e., $1301 \bmod 27 = 5$. She proceeds to translate 5 to row $i = 6$ for the next query and adds the solution of row $i = 6$, i.e., 256391478, to the proof string. Last, she attaches the hashes of the Merkle tree nodes needed to check whether the result matches the commitment to $\mathcal{P}$. They are illustrated in the green boxes in Figure 6A. The proof string now is

$$\mathcal{P} = eb47\ 914783256\ e7f1\ bdb1\ 185f\ 0248\ 0610\ 256391478\ 7704\ 7cc9\ 747d\ 2176\ 43c6.$$

Peggy could perform additional querying rounds, but we stop here to not blow up the proof size. She propagates her proof to the Sudoku community such that any verifier can now check the proof by only using the proof string and the initial Sudoku puzzle.

Next we describe how a verifier named Victor checks the proof. He is aware that Peggy queried twice, and since he knows the protocol, he can derive exactly which characters in the proof string have which meaning;

i.e., the first 4 characters represent the Merkle root, the subsequent 9 represent a solution for a row, column, or box, the next 20 characters are five hashes used in the Merkle tree, etc. Victor then verifies the proof as follows:

1. He reads the first four characters in the proof string that correspond to the Merkle root. He calculates which row, column, or box is queried first by transforming the Merkle root to a decimal number in $\mathbb{Z}_{27}$ as described above. He finds that the first query concerns box $b = 4$.
2. From the proof string, he reads the characters 5 to 13, which correspond to the solution in row $i = 5$ and then hashes it; i.e., $H(914783256) = 0515$.
3. He reads the subsequent 20 characters from the proof string and splits them up in five hashes with four characters each; i.e., $h_1 = \{e7f1, bdb1, 185f, 0248, 0610\}$. He then uses the solution's hash above—0515—and hashes it pairwise with the hashes in h_1. He can then check whether this matches the Merkle root, i.e., the first four characters in the proof string.
4. He uses the solution's hash in row $b = 4$ and transforms it to a decimal number in $\mathbb{Z}_{27}$. He will find that the next query is row $i = 6$.
5. From the proof string, he reads the characters 34 to 42, which correspond to the solution in column $i = 6$ and then hashes it; i.e., $H(256391478) = 5f19$.
6. He reads the last 20 characters from the proof string and splits them up in five hashes with four characters each; i.e., $h_2 = \{7704, 7cc9, 747d, 2176, 43c6\}$. He then uses the solution's hash above—$5f19$—and hashes it pairwise with the hashes in h_2. He can then check whether this matches the Merkle root, i.e., the first four characters in the proof string.

Any other verifier can replicate these steps to verify Peggy's proof. Obviously, there are many caveats in this still-easy example. First, the proof is not zero knowledge. With each additional round, Peggy reveals new information on the solution. To keep it zero knowledge, she would need to apply more complex mathematical methods, which would go beyond the scope of this article (see Berentsen, Lenzi, and Nyffenegger, 2022).

Second, the proof is not succinct as the proof string $\mathcal{P}$ has 62 characters. Since this example is very simple, propagating the digits in all fields of the puzzle would only contain either 81 or only 59 characters, depending on whether the initial values are included or not. However, for larger Sudoku puzzles, succinctness becomes relevant. A puzzle of size 25×25 with 119 initial values would require Peggy to send 506 characters if she sends the whole solution. Applying the protocol from above, she needs 4 characters for the Merkle root and in each round 25 characters for the solution and $7 \times 4 = 28$ for the hashes in the Merkle tree. This yields a proof size of $4 + 53n$, where n is the number of querying rounds. The proof size in this example is smaller than just sending the whole solution for $n < 10$.

It is possible to transform most interactive proofs into non-interactive ones. However, as seen above, the transformation makes the proof more complex. Interactive proofs are often simpler and more efficient than non-interactive ones (ZKProof, 2022). Since non-interactive proofs can be verified by any party at any time, they still have broad use cases as in, for example, blockchain applications.

3.3 Different Proof Protocols

When it comes to real-world applications, several different proof protocols can be used. Non-interactive proofs are the ones mostly used, and the generic term is often denoted as a SNARK, which was first introduced by Bitansky et al., 2012. SNARK stands for **S**uccinct **N**on-Interactive **AR**gument of **K**nowledge. "Succinct" means that verifying a proof is so much more efficient than reexecuting the computation and non-interactive that a prover can provide a single proving object that can be verified by everybody.

To understand the definition of "argument," we need to clarify the distinction between a proof and an argument. A proof needs to be sound (see Section 3, p. 283) even against a computationally unbounded malicious prover.[13] On the other hand, an argument relies on a weaker assumption and only requires soundness against a computationally bounded malicious prover (ZKProof, 2022). A computationally unbounded malicious prover could, for example, find collisions in cryptographic hash functions, which would be a serious security concern.[14] "Of knowledge" means that the prover must also prove that the secret input data are known (Szepieniec, 2022).

There are several different proof protocols in the line of a SNARK such as Groth16 (Groth, 2016), PlonK (Gabizon, Williamson, and Ciobotaru, 2019), Merlin (Chiesa et al., 2019), Bulletproofs (Bünz et al., 2018), or Nova (Kothapalli, Setty, and Tzialla, 2021). Many SNARKs rely on elliptic curves[15] and thus on the assumption

13. The term "computationally unbounded" refers to a theoretical concept in which an entity has such a huge amount of computing power that any computation can be executed rapidly. In practice there are hardware limitations that make certain theoretical attacks unfeasible.

14. Finding collisions of the mostly implemented cryptographic hash functions requires an enormous amount of computational power. In practice, a computationally bounded attacker cannot successfully perform this attack with a very high probability.

15. Also, the private-public key encryption used to access crypto assets in a wallet normally relies on elliptic curves.

that discrete logarithms are difficult to compute (Thaler, 2022). However, this assumption is not postquantum secure.[16]

Furthermore, using elliptic curves in a ZKP requires that certain parameters are set up in advance. These parameters are then used when creating a proof. The problem is that if somebody knows how these parameters were created, he could cheat and create false proofs. As a result, these parameters are normally created in initial set-ups called ceremonies, in which several parties participate. All parties contribute to the set-up by secretly creating random elements that are then combined to get the parameters. If it is somehow possible to get hold of all random elements, it would be possible to create false proofs. Thus, these elements are often called toxic waste. However, if at least one party in the ceremony destroys the personal random element, it is not feasible to derive the composition of the parameters and the proofs are secure.

Due to these shortcomings, there has been an effort to create ZKPs that do not rely on elliptic curves. The most prominent protocol that works without elliptic curves are STARK proofs, introduced by Ben-Sasson et al., 2018. The main cryptographic technique used are cryptographic hash functions. Hence, STARK proofs are postquantum secure and do not need an initial set-up or ceremony, and therefore a S**T**ARK is called **T**ransparent. Furthermore, the **S** in **S**TARK does not stand for succinct but for **S**calable. In addition to having efficient verification, a STARK entails some efficiency assumptions on the creation of the proof by the prover (Szepieniec, 2022).

An advantage of a SNARK over a STARK proof is that the proof size is considerably smaller. This becomes especially relevant when it comes to blockchains in which block space is costly and thus users are interested in shorter proofs. Furthermore, in STARKs there is an inherent trade-off between verification costs and security that is not prevalent in SNARKs, which rely on elliptic curves (Thaler, 2022).

4. APPLICATIONS IN BLOCKCHAIN

When it comes to blockchains, there are two main applications of ZKPs: data confidentiality and efficiency. An example for the former use case is privacy-protecting payments, whereas for the latter, it is increased scalability by implementing rollups.

4.1 Data Confidentiality

An example where data confidentiality is applied are privacy-protecting transfers. In standard blockchains, each transaction is registered on the blockchain and is hence visible to everybody. Even though a user remains pseudonymous, it can be possible to derive a real identity using transaction data from and to an address. By using ZKPs, it is possible to prove that someone is allowed to spend a certain amount without revealing the sender and receiver address and amount to the whole network.

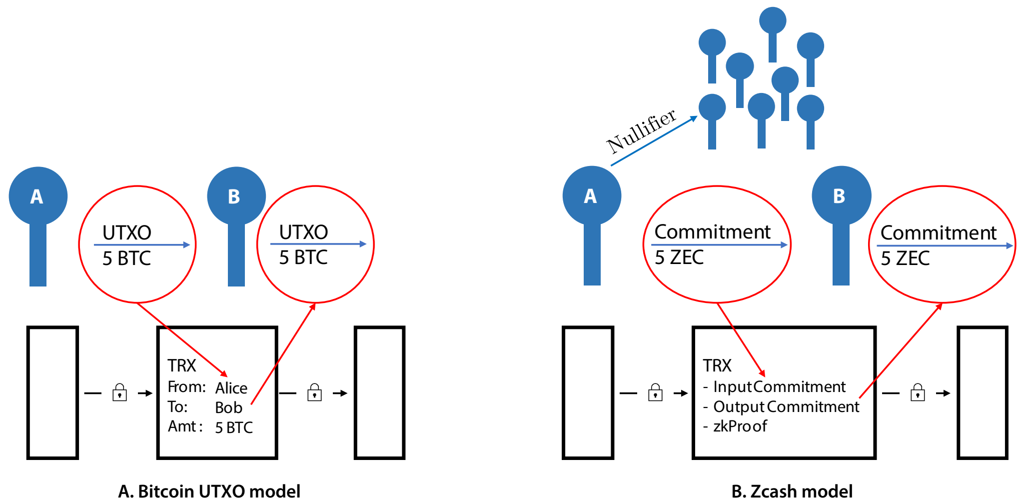
Zcash, introduced by Ben Sasson et al., 2014, is probably the most famous cryptocurrency that leverages ZKPs for private transactions. To understand how Zcash works, we first recap Bitcoin's (BTC) transaction model. The Bitcoin network uses unspent transaction outputs (UTXOs) to determine which address is allowed to spend a specific amount of the currency, illustrated in Figure 7A. Consider an example where Alice receives 5 Bitcoin (BTC) in a transaction from another network participant. The network stores the information that a UTXO worth 5 BTC exists. If Alice wants to send 5 BTC to her friend Bob, she creates a transaction in which she references the UTXO. She also needs to sign the transaction with her private key to prove that she can rightfully spend the UTXO. Next, Alice propagates the transaction to the network, which then includes it in a block. By sending the 5 BTC to Bob, a new UTXO is created that he can use again.

In Zcash a similar construct is used called commitment. Say that Alice has received funds via the Zcash blockchain, which comes in the form of a commitment that, contrary to a UTXO, does not reveal the address and the available amount to spend to the network. Conceptually, it is a hash of her address, the spendable amount, and a serial number denoted as s; i.e., $Commitment = H(address, amount, s)$. If Alice creates a transaction to send the funds to Bob, she does the following (see Figure 7B):

1. She prepares a ZKP to prove that she is allowed to spend the input commitment that corresponds to the commitment's hash.
2. She creates a new output commitment, which is a hash of Bob's address, the amount, and the serial number s.
3. She propagates the commitment together with her ZKP to the network, which then includes it in a block.

16. A quantum computer can very efficiently factor integers using Shor's algorithm, which can be used to break elliptic curve cryptography. However, current quantum computers are far from having enough computational power to break the cryptography in a reasonable time see, e.g., Webber et al., 2022.

Figure 7
Difference between the UTXO model in Bitcoin and the Zcash Model



4. Simultaneously, she creates a nullifier that is derived from the serial number *s*. She propagates it to the network, which stores all nullifiers. The nullifier set is a list of hashes that show which input commitments have been spent. If she wanted to double spend the same input commitment again, the same nullifier would result and the transaction would be considered invalid by the network.
5. She privately sends Bob the information that the output commitment relates to his address, the amount included, and the serial number *s*. Bob needs this information to be able to spend the funds himself.

The ZKP proves she is in possession of the private key corresponding to the address and that the amount of the input commitment equals the amount of the output commitment without revealing her address nor the amount.

4.2 Efficiency

A trade-off that many blockchain networks face is the trilemma of security, decentralization, and scalability. Many protocols that emphasize their role of a decentralized network, such as, for example, Bitcoin and Ethereum, face at some point the problem that the demand for block space is much higher than the supply. In other words, the demand for sending transactions exceeds what the network can handle, which in turn makes sending a transaction in these networks pretty expensive.

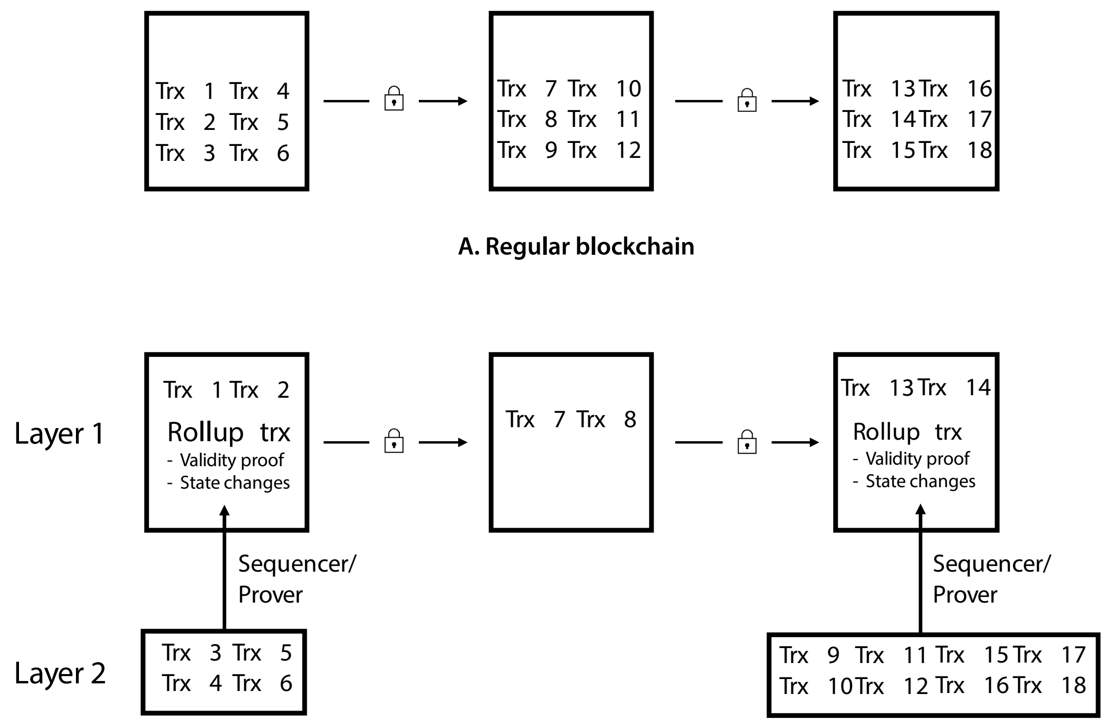
It would be relatively easy to increase scalability by raising the number of transactions that can be included in a block. However, a highly decentralized network relies on the assumption that any regular person who is interested in participating in the network and verifying that the state of the network is correct can do so with a reasonable hardware (e.g., with a regular desktop computer). But if more transactions are sent over the network, the network needs to store more data and the verification of all the transactions becomes more expensive, which in turn requires better hardware and excludes some participants from verifying the state of the network. Thus, blockchain developers are interested in finding solutions that increase scalability without harming decentralization or security.

The use of ZKPs promises to remedy exactly that through the use of so called rollups, which are a layer 2 scaling solution.[17] In a regular blockchain all transactions are included within blocks of the base chain as illustrated in Figure 8A. As discussed earlier, verifying all transactions can be very expensive—a big computational burden is to check all the signatures to ensure that someone is allowed to execute a certain transaction. The idea of a rollup is that a network participant called a sequencer[18] batches (or rolls up) transactions off-chain (on layer 2) and creates a proof that proves the validity of these transactions on layer 1. This process is depicted in Figure 8B.

17. Layer 1 would be the base chain (e.g., Ethereum). A layer one blockchain can be scaled directly by, for example, increasing the block size. Rollups scale a layer 1 blockchain indirectly and hence the scaling is referred to as a layer 2 scaling solution.

18. In theory, the sequencer could be decentralized or a centralized party, but in practice, sequencers are currently centralized in real implementations. However, there is work ongoing to decentralize the sequencer as, for example, in the case of the StarkNet rollup (see, e.g., StarkWare, 2022).

Figure 8
Rollups Illustrated Conceptually



Consequently, network participants do not have to verify each individual transaction and signature but only the compressed proof, which is much less expensive. The proofs are based on ZKPs but currently do not incorporate privacy and instead focus on proving the validity of the transactions in a succinct way. Thus in this context, they are often called validity proofs and not ZKPs.[19] Basically, what happens is that on layer 1 (on the main chain), a smart contract is created that acts as the proof's on-chain verifier. The sequencer is the prover who creates proofs off-chain on layer 2 and sends a transaction to the on-chain smart contract, which verifies the proof.

So far, we only described the validity of the transactions. However, we also need to record in the blockchain who sent a transfer to whom. This recording is done by publishing the state changes[20] in the call data[21] of the transaction sent to the verifier smart contract. In that way, the current account balances are always updated on-chain. Hence, executing (i.e., checking the validity of the transactions and signatures) is done off-chain, but settling the transactions (i.e., state changes) is on-chain.[22] By using a rollup, much less block space is used on the

19. Note that there is another widely used type of rollup, i.e., optimistic rollups. Optimistic rollups do not rely on ZKPs, which is why we do not discuss them here. They are optimistic in the sense that any network participant can propose a state change on-chain. Afterwards, some network participants check the state change, and if a transaction in it is invalid, a fraud proof can be posted on-chain, which reverts the state back. Thus, optimistic proofs have a period of about one week until a transaction is final, in which network participants have time to check the state change. The difference between ZKP rollups and optimistic rollups may become more intuitive in the following joke: Two rollups walk into a bar. The barman asks, "Can I see your ID's?" The ZKP rollup says, "I can prove to you I'm over 18, but I won't show you my ID." The optimistic rollup says, "If nobody can prove I'm underage in 7 days, that means I'm over 18."

20. Ethereum works with an account-based model. At each period in time, each account is assigned a specific balance, which can be interpreted as a huge spreadsheet and is called the state. The state is hashed in a tree, and the root is written into each newly produced block. If a transaction is sent from Alice to Bob, the balance of Alice decreases, whereas the one of Bob increases; this process is called a state change. After the transaction, there is a new state, which is again published to the blockchain.

21. The call data are an additional field in a transaction in which some arbitrary data can be written. Note that writing data into the call data costs gas (i.e., the native way to pay for the inclusion of a transaction in the Ethereum network). However, there is a discussion in the Ethereum network to reduce the call data gas cost to make it more rollup-friendly (Ethereum Improvement Proposal 4488).

22. By now, there are many different types of rollups depending on where the data are stored, who performs the execution, etc. See,

base chain. Additionally, since the computational effort used to verify the proof scales logarithmically with the size of the input data (i.e., the number of transactions), the fee per transaction when using a rollup decreases in the number of transactions in the rollup and is generally much lower compared to sending a transaction directly on layer 1.

Note that it is still possible to make regular transactions on layer 1, as illustrated in Figure 8B, where some transactions are batched on layer 2 but some are still on the main chain. Moreover, a rollup does not need to send a transaction batch each block. The sequencer could wait until enough transactions are accumulated and only publish a batch every few blocks and in doing so reduce the fee per transaction. This process, however, has the disadvantage of a delayed finality time, and if the sequencer is not decentralized, it comes with some counterparty risk.

On layer 2, developers can build any type of smart contract or application. It is even possible to think of a layer 3 or layer 4 to further increase scalability. In each layer transactions would be batched and sent to the lower layer. However, the more layers that are introduced, the longer the finality time for the upper layers.

There are also other blockchain applications that use ZKPs for efficiency reasons. For example, Filecoin uses them for a decentralized storage application. Instead of storing data with a large company providing a cloud service, Filecoin enables storing data in a decentralized way. To do that in a confidential way, the data to be stored are split up into small chunks and are stored by several nodes across the network.[23] The storage providers earn a reward by storing the data, and they use a ZKP to efficiently prove that they do store them correctly.

5. APPLICATION IN OTHER AREAS

ZKPs potentially have many other use cases apart from blockchains. Even though the theoretical concepts have been around for some time, ZKPs have not been widely adopted yet. There are many possible applications with regard to data confidentiality—this includes privacy and identity management—or efficiency. A basic requirement to use ZKPs in real-world problems is that for each application there must be some underlying cryptographic set-up. When it comes to blockchains, this is naturally existent. However, in real-world examples we would often need to rely on some trusted third parties, as we show below.

5.1 Data Confidentiality

Being private about personal characteristics is one example where ZKPs could be used for data confidentiality. As an example, Carlos wants to prove at the entrance of a nightclub that he is 18 without revealing his exact age or any other personal characteristic. Judy wants to enter a fancy club she has a membership to but does not want the doorman to know who she is. Both could create a ZKP to prove that they are allowed to enter. For both examples, we would need some kind of underlying cryptographic set-up. In the former case, the government might have provided a digital identity, whereas in the latter case the club owners could have issued a digital membership card that they signed. The doorman could verify that the proof includes the signature of the club owner and would let Judy in.

ZKPs could also be used in areas in economics that are not directly linked to blockchains. Assume the company funInvest wants to prove that it is solvent without revealing its exact portfolio composition. It is not straightforward to use a ZKP in this environment since an underlying cryptographic set-up with respect to their asset holdings is missing. If, for example, funInvest owns some shares of company X, a trusted third party, the company X itself could instead sign a message that funInvest owns shares above a certain amount without revealing the exact amount. This becomes easier when funInvest holds some crypto asset. It could then prove that it holds the private key to an address that has an amount of assets above a certain threshold, without revealing the exact amount or the address.[24]

A proposal from ING, 2017 goes into a similar direction. For example, ZKPs could be used for a mortgage applicant to prove that her salary is in a specific range needed to receive the mortgage without revealing her exact salary.

5.2 Efficiency

An example where ZKPs increase efficiency is when you can create a short ZKP instead of filing a ton of paper work. ZKProof, 2022 discusses one example about airplanes. Airplanes undergo a lot of checks and maintenance, which a regulatory body checks. Anyone who works on the plane files some papers, which

---

e.g., Charbonneau, 2022 for an overview.

23. This storing is done with some redundancy such that even if some storage providers fail, the data are still available.

24. Even though the verifier learns that funInvest holds funds above the threshold, the zero knowledge property still holds. Zero knowledge means that the verifier cannot learn more than the statement to be proven and the statement is explicitly about the threshold.

the airline then collects and forwards to the regulatory body. Checking all the paperwork is quite inefficient. Instead, the regulatory body could define which workers are trusted to make the checks. After conducting the checks, the workers can cryptographically sign that they did the work instead of filing papers and provide a digital receipt to the airline. The airline could then combine all the digital receipts and provide them to the regulatory body, which proves that the workers did indeed do the checks and maintenance work. The proof can then be efficiently verified by the regulatory body.

6. CONCLUSION

In this article we discussed ZKPs and their applications. ZKPs can make certain digital processes considerably more efficient and can be used in privacy applications. There are a lot of use cases when it comes to blockchains but also in broader areas of economics. For blockchains, ZKPs can be used to improve scalability or to protect privacy, such as when using the Zcash cryptocurrency.

With this review article, we aim to shed light on the topic and provide a basic understanding to a broad audience. For the more ambitious reader, Berentsen, Lenzi, and Nyffenegger, 2022 go through the math of an easy ZKP and provide an accompanying Python script.

REFERENCES

Ben Sasson, Eli, Alessandro Chiesa, Christina Garman, Matthew Green, Ian Miers, Eran Tromer, and Madars Virza. 2014. Zerocash: decentralized anonymous payments from bitcoin. *2014 IEEE Symposium on Security and Privacy,* 459–474. https://doi.org/10.1109/SP.2014.36.

Ben-Sasson, Eli, Iddo Bentov, Yinon Horesh, and Michael Riabzev. 2018. Scalable, transparent, and post-quantum secure computational integrity. *International Association for Cryptologic Research,* ISSN: 1746-8094. https://eprint.iacr.org/2018/046.pdf.

Berentsen, Aleksander, Jeremias Lenzi, and Remo Nyffenegger. 2022. *A walk-through of a simple zk-stark proof.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4308637.

Bitansky, Nir, Ran Canetti, Alessandro Chiesa, and Eran Tromer. 2012. From extractable collision resistance to succinct non-interactive arguments of knowledge, and back again. *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference on - ITCS '12.,* 326–349.

Bünz, Benedikt, Jonathan Bootle, Dan Boneh, Andrew Poelstra, Pieter Wuille, and Greg Maxwell. 2018. Bulletproofs: short proofs for confidential transactions and more. *2018 IEEE Symposium on Security and Privacy (SP).*

Burger, Elena, Bryan Chiang, Sonal Chokshi, Eddy Lazzarin, Justin Thaler, and Yahya Ali. 2022. Zero knowledge canon, part 1 & 2, September. Accessed September 21, 2022. https://a16zcrypto.com/zero-knowledge-canon/.

Buterin, Vitalik. 2016. Quadratic arithmetic programs: from zero to hero. Accessed April 19, 2022. https://medium.com/@VitalikButerin/quadratic-arithmetic-programs-from-zero-to-hero-f6d558cea649.

Chalkias, Konstantinos, and Mike Hearn. 2019. Demonstrate how zero-knowledge proofs work without using maths. *CORDACON 2017 Conference* (February).

Charbonneau, Jon. 2022. The complete guide to rollups. Accessed September 22, 2022. https://members.delphidigital.io/reports/the-complete-guide-to-rollups.

Chiesa, Alessandro, Yuncong Hu, Mary Maller, Pratyush Mishra, Psi Vesely, and Nicholas Ward. 2019. Marlin: preprocessing zksnarks with universal and updatable srs. *Cryptology ePrint Archive, Paper 2019/1047.*

Gabizon, Ariel, Zachary J. Williamson, and Oana Ciobotaru. 2019. Plonk: permutations over lagrange-bases for oecumenical noninteractive arguments of knowledge. *Cryptology ePrint Archive, Paper 2019/953.*

Goldwasser, Shafi, Silvio Micali, and Charles Rackoff. 1989. The knowledge complexity of interactive proof systems. *SIAM Journal on Computing* 18 (1): 186–208.

Gradwohl, Ronen, Moni Naor, Benny Pinkas, and Guy N. Rothblum. 2009. Cryptographic and physical zero-knowledge proof systems for solutions of sudoku puzzles. *Theory of Computing Systems* 44 (2): 245–268. ISSN: 14324350. https://doi.org/10.1007/s00224-008-9119-9.

Groth, Jens. 2016. On the size of pairing-based non-interactive arguments. *Advances in Cryptology – EUROCRYPT 2016,* https://doi.org/10.1007/978-3-662-49896-5.

ING. 2017. Ing launches zero-knowledge range proof solution, a major addition to blockchain technology. Accessed April 16, 2022. https://www.ingwb.com/en/insights/distributed-ledger-technology/ing-launches-major-addition-to-blockchain-technology.

Kothapalli, Abhiram, Srinath Setty, and Ioanna Tzialla. 2021. Nova: recursive zero-knowledge arguments from folding schemes. *Cryptology ePrint Archive, Paper 2021/370.*

Nadler, Matthias, and Fabian Schaer. 2023. Tornado cash and blockchain privacy: a primer for economists and policy makers. *Federal Reserve Bank of St. Louis Review* 105 (2).

Naor, Moni, Yeal Naor, and Omer Reingold. 1999. Applied kid cryptography or how to convince your children you are not cheating. *Journal of Cryptology.*

Schaer, Fabian, and Aleksander Berentsen. 2020. *Bitcoin, blockchain, and cryptoassets: a comprehensive introduction.* MIT Press, September.

StarkWare. 2022. Starknet: on to the next challenge. Accessed May 13, 2022. https://medium.com/starkware/starknet-on-to-the-next-challenge-96a39de7717.

Szepieniec, Alan. 2022. Anatomy of a stark, part 1: stark overview. Accessed September 22, 2022. https://aszepieniec.github.io/stark-anatomy/overview.

Thaler, Justin. 2022. Snark security and performance. Accessed September 22, 2022. https://a16zcrypto.com/snark-security-and-performance/.

Webber, Mark, Vincent Elfving, Sebastian Weidt, and Winfried K. Hensinger. 2022. The impact of hardware specifications on reaching quantum advantage in the fault tolerant regime. *AVS Quantum Science* 4 (1): 013801. ISSN: 26390213. https://doi.org/10.1116/5.0073075.

ZKProof. 2022. Zkproof community reference (version 0.3), https://docs.zkproof.org/pages/reference/versions/ZkpComRef-0-3.pdf.