ETH zürich

KOF

KOF Dissertation Series

Empirical Shock Analysis

Marc Anderes

Diss. ETH No. 29266

KOF Dissertation Series, No. 54, 2023

Imprint

Publisher

KOF Swiss Economic Institute, ETH Zurich

DISS. ETH NO. 29266

Empirical Shock Analysis

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

MARC ANDERES

M.Sc. in International and Monetary Economics, University of Bern

born on 20.01.1990

accepted on the recommendation of

Prof. Dr. Jan-Egbert Sturm, examiner

Prof. Dr. Stefan Pichler, co-examiner

2023

Contents

List of Figures

List of Tables

Abstract

This dissertation is a collection of articles on the empirical analysis of shocks. Each research article is dedicated to a different type of shock, and so the statistical methods employed in this dissertation build on both macro- and microeconometrics. The dissertation begins with an introduction, describing the structure and discussing the contribution to the existing literature. It is followed by four Chapters containing the research articles. In the case of co-authored Chapters, all authors have contributed equally to the work.

The first Chapter estimates the dynamic effects of housing demand shocks on a large set of macroeconomic series and detailed household balance sheet components for different wealth groups. Results show that a positive housing demand shock translates into a large boom in economic activity and reveal notable heterogeneity among wealth groups. While households of all wealth groups make heavy use of home equity-based borrowing, I find a larger consumer spending sensitivity for poorer households. A historical decomposition suggests that housing demand shocks have largely contributed to the pronounced drop in poorer households' consumption during and after the Great Recession.

The second Chapter, coauthored with Florian Eckert and Nina Mühlebach, focuses on output gap reconciliation. We use a state-space framework to estimate a 'true' output gap that is reconciled from ten well-established output gap methods. The use of multiple vintages allows us to model revisions and decompose them into news and noise components, leading to an economically meaningful output gap estimate. In a comprehensive real-time study for the United States, we review the performance of all output gap methods and provide evidence that a) our reconciled

gap is a reliable real-time estimate and b) that the data prefer the inclusion of multiple releases and methods.

In the third Chapter, coauthored with Alexander Rathke, Sina Streicher, and Jan-Egbert Sturm, we examine whether ECB communication shocks add information to a shadow interest rate that summarizes the overall policy stance as interpreted by financial markets. To measure communication, we use information based on ECB press releases distinguishing between topics like inflation, the real economy and monetary developments. We also look at the effect of communication on consensus expectations about key macroeconomic variables. The ECB's assessment of the economy, i.e., communication related to economic growth, triggers movements in financial markets and, therefore, the shadow rate. Communication of the ECB through its press releases also causes professional forecasters to change their outlooks. Not only are their growth forecasts affected, but so are their expectations for M3 growth and inflation.

The fourth Chapter, coauthored with Stefan Pichler, looks at the mental health effects of the COVID-19 pandemic shock. We combine data from the most popular Swiss help hotline with administrative data from the phone network provider to study how COVID-19 and related social distancing measures affected mental health. Our results show a significant increase in the number (+10%) and duration (+5%) of calls during the first wave. An analysis of unanswered calls shows an unmet demand for counseling during the first wave. Men and people over 65 years of age showed the largest increase in call volume and duration. Finally, we find small and sometimes negative effects during the second wave, where Switzerland was among the least restrictive countries in terms of social distancing.

Zusammenfassung

Diese Dissertation ist eine Sammlung von Forschungsartikeln über die empirische Analyse unterschiedlicher Typen von Schocks. Die methodische Herangehensweise ist jeweils abhängig von der Natur des betrachteten Schocks, wobei makro- und mikroökonometrische Methoden verwendet werden. Die Dissertation beginnt mit einer Einleitung, in der die Struktur der vorliegenden Arbeit und der Beitrag zur bestehenden Literatur diskutiert werden. In den darauffolgenden vier Kapiteln folgen die Forschungspapiere. Bei gemeinsam verfassten Kapiteln haben alle Autoren gleichermassen zur Arbeit beigetragen.

Im ersten Kapitel werden die dynamischen Auswirkungen von Schocks bei der Immobiliennachfrage auf eine Vielzahl makroökonomischer Reihen und detaillierter Bilanzkomponenten der privaten Haushalte für verschiedene Vermögensgruppen geschätzt. Die Ergebnisse zeigen, dass ein positiver Schock bei der Wohnungsnachfrage zu einem Boom der Wirtschaftstätigkeit führt, und lassen eine bemerkenswerte Heterogenität zwischen den Vermögensgruppen erkennen. Während Haushalte aller Vermögensgruppen in hohem Masse auf Eigenheimkredite zurückgreifen, stelle ich fest, dass ärmere Haushalte stärker auf Konsumausgaben reagieren. Eine historische Zerlegung deutet darauf hin, dass Schocks bei der Immobiliennachfrage in hohem Masse zum ausgeprägten Rückgang des Konsums der ärmeren Haushalte während und nach der Grossen Rezession beigetragen haben.

Das zweite Kapitel, welches in Zusammenarbeit mit Florian Eckert und Nina Mühlebach entstand, fokussiert sich auf das Schätzen von Produktionslücken. Wir verwenden ein State-Space Modell, um aus zehn etablierten Methoden eine "wahre" Produktionslücke zu rekonzilieren. Die Verwendung mehrerer Echtzeit-Datenreihen (sog. "vintages") ermöglicht es uns, Revisionen zu modellieren und sie in News- und Noise-Komponenten zu zerlegen, was zu einer wirtschaftlich aussagekräftigen Produktionslückenschätzung führt. In einer umfassenden Echtzeit-Studie für die Vereinigten Staaten überprüfen wir das Abschneiden aller Methoden und weisen nach, dass a) unsere rekonzilierte Lücke eine zuverlässige Echtzeit-Schätzung ist und b) dass die Daten die Einbeziehung mehrerer Veröffentlichungen und Methoden vorziehen.

Im dritten Kapitel, das gemeinsam mit Alexander Rathke, Sina Streicher und Jan-Egbert Sturm verfasst wurde, untersuchen wir, ob Kommunikationsschocks der europäischen Zentralbank (EZB) Informationen zu einer "shadow rate" hinzufügen, welche den allgemeinen politischen Kurs, wie er von den Finanzmärkten interpretiert wird, zusammenfasst. Zur Messung der Kommunikation verwenden wir Informationen aus EZB-Pressemitteilungen, die zwischen Themen wie Inflation, Realwirtschaft und monetären Entwicklungen unterscheiden. Ausserdem untersuchen wir die Auswirkungen der Kommunikation auf die Erwartungen professioneller Prognostiker hinsichtlich der wichtigsten makroökonomischen Variablen. Die Einschätzung der Wirtschaft durch die EZB löst Bewegungen auf den Finanzmärkten und somit auch der "shadow rate" aus, auch wenn man für die erwartete und tatsächliche Wirtschaftsentwicklung kontrolliert. Die Kommunikation der EZB über ihre Pressemitteilungen führt zudem dazu, dass Prognostiker ihre Prognosen revidieren. Davon sind nicht nur ihre Wachstumsprognosen betroffen, sondern auch Erwartungen für das M3-Wachstum und die Inflation.

Das vierte Kapitel, das gemeinsam mit Stefan Pichler verfasst wurde, befasst sich mit den Auswirkungen des COVID-19-Pandemieschocks auf die psychische Gesundheit. Wir kombinieren Daten der beliebtesten Schweizer Helpline mit administrativen Daten des Telefonnetzbetreibers, um zu untersuchen, wie sich COVID-19 und die damit verbundenen Massnahmen zur sozialen Distanzierung auf die psychische Gesundheit auswirken. Unsere Ergebnisse zeigen einen signifikanten Anstieg der Anzahl (+10%) und Dauer (+5%) der Anrufe während der ersten Welle. Eine Analyse der unbeantworteten Anrufe zeigt, dass die Nachfrage nach Seelsorge während der ersten Welle nicht gedeckt wurde. Männer und Personen über 65 Jahre verzeichneten den grössten Anstieg des Anrufvolumens und der Anrufdauer. Schliesslich stellen wir in der zweiten Welle geringe und teilweise negative Auswirkungen fest, wobei die Schweiz zu den am wenigsten restriktiven Ländern in Bezug auf die soziale Distanzierung gehörte.

Introduction

Empirical economics has undergone significant changes in the last decades. As Edward Leamer (1983) famously observed: "Hardly anyone takes data analysis seriously. Or perhaps more accurately, hardly anyone takes anyone else's data analysis seriously." Fortunately, this is no longer true. In a recent response to Leamer's original article, Angrist and Pischke (2010) note that "empirical economics has experienced a credibility revolution", which has been caused by improved research designs, advances in theoretical econometric understanding, and the availability of more and better data. In this thesis, my co-authors and I empirically investigate how different types of shocks influence the lives of households, markets, or individuals.

In the context of economics, a shock is an unexpected event that disrupts the normal functioning of the economy, often leading to changes in economic variables such as gross domestic product (GDP), inflation, interest rates, and employment. Examples of economic shocks include natural disasters, policy changes, oil supply disruptions, and pandemics. Since shocks are by definition unpredictable, they can in certain circumstances provide a useful tool for inferring a causal mechanism. For example, a change in tax policy or an unexpected shift in monetary policy could be regarded as a natural experiment. By comparing outcomes before and after the shock, researchers may be able to identify the causal effects of such a policy change. Another way to infer causality is when a group is exposed to a shock, while another is not. Comparing the difference in changes between the two groups may then reveal the causal effect of the shock.

Understanding the causal relationships between economic shocks and other variables is especially important for policymakers. It allows them to mitigate the negative effects of shocks, while also designing policies that create economic and social opportunities. However, using shocks to infer causality requires careful consideration of the underlying mechanisms and context of the shock, as well as the appropriate methods. Depending on the question at hand, empirical economists frequently use tools from either micro- or macroeconometrics. Macroeconometrics is concerned with modeling the behavior of the economy as a whole, typically at the level of aggregates such as GDP or inflation. This kind of data often implies a strong reliance on time-series analysis. On the other hand, microeconometrics is focused on analyzing the behavior of individual economic agents like households or firms. This branch of econometrics thus tends to rely more heavily on panel data techniques, which track the economic agents over time. Since my co-authors and I analyze a variety of shocks on different levels of analysis, we employ methods from both subfields of econometrics. The remainder of this introduction is structured in two parts. First, I will discuss the different econometric methodologies and how they are related. The second part explains the nature of the various shocks that are analyzed, the literature surrounding them, and how we contribute to it.

The first two chapters use Bayesian state-space models to isolate either housing demand shocks (Chapter 1) or news and noise measurement shocks (Chapter 2). State-space models are powerful statistical tools in time-series analysis that describe relationships between possibly unobserved processes and known measurements (see Chan and Strachan, 2023, for a detailed overview). Notable examples of unobserved processes that can be estimated in this framework are time-varying parameters in regression models, stochastic volatility models, and trend-cycle decompositions. Another prominent example is given by the application in Chapter 1, in which I extract latent dynamic factors from a high-dimensional set of time-series. In Chapter 2, my co-authors and I decompose the revisions of output gap estimates into unobserved news and noise components. Both chapters rely on Bayesian methods to estimate the relatively complex state-space model. The advantage of this procedure is that the likelihood-based approach is able to explicitly exploit the structure of the state equation in the estimation of the factors (see Bernanke et al., 2005). As a byproduct, it also delivers full posterior distributions naturally quantifying the extent of uncertainty. Both chapters use a multimove Gibbs sampler (Carter and Kohn, 1994; see Kim and Nelson, 1999 for an overview) in which the parameters are sampled conditional on the most recent draws of the hidden states, and the hidden states are drawn conditional on the most recent parameter estimates.

The third and fourth Chapter use an event-study design to analyze the effects of communication shocks (Chapter 3) and of a pandemic shock (Chapter 4). An event-study examines the behavior of agents or, for example, financial variables around an event and tries to measure the impact of the shock. The main part in Chapter 3 does so in a high-frequency environment, as we study the influence of the European Central Bank's (ECB) communication on the shadow interest rate, which proxies the stance of monetary policy. We analyze changes in the interest rate within a narrow one-day window around the ECB's press release dates. The advantage of this methodology is that it holds the macroeconomic outlook constant so that the effects of communication shocks can be isolated (see also the discussion in Swanson, 2011). The time window in the event study in Chapter 4, however, is much wider, as we analyze the developments in the months after the outbreak of the COVID-19 pandemic with the preceding years.

In summary, the unifying element of the chapters in this dissertation is the empirical investigation of shocks using a rich set of econometric tools. My co-authors and I explore and evaluate four shocks that are very different in their nature. In the first Chapter, I examine the effects of housing demand shocks on a high-dimensional set of U.S. macroeconomic series and detailed household balance sheet components for four wealth percentile groups. By doing so, the paper combines different strands of literature. First, the financial crisis showed that housing shocks need to be considered in order to understand business cycle fluctuations (Iacoviello and Neri, 2010; Liu et al., 2013). Second, the strength of housing shocks depends on the heterogeneity of households' balance sheets (Mian et al., 2013), since housing wealth is more important for households in lower wealth percentiles in determining borrowing constraints, and thus, consumption (Berger et al., 2017). Third, to model the rich joint dynamics of the real economy, the financial sector, and households' balance sheets, I use a dynamic factor model as an econometric framework. The paper's main contributions are the following. First, the identification of the housing demand shock is arguably sharper than in the literature. The reason is that the high-dimensional model environment allows me to impose a large number of theoretically motivated restrictions, for example on house prices and building permits in different regions. Second, building on the identified housing shocks, I analyze the consumption response of four different wealth percentile groups. The results show that the reaction of durable consumption is a decreasing function of net wealth, while variance decompositions imply that the response of the poorest 50% of households is largely explained by the identified housing shocks. Finally, I compute elasticities of consumption with respect to house prices, which suggest a larger spending sensitivity for weaker balance sheet households.

In the second Chapter, my co-authors and I use a state-space framework to reconcile output gaps from various methods. We use multiple vintages to decompose revisions into news and noise shocks, while also computing a reconciled output gap estimate. The output gap is a measure of the cyclical position of the economy, and thus an important part of central banks' reaction functions (Sturm and De Haan, 2011). It is also used to determine structural budget balances for fiscal planning or in financial regulation frameworks to estimate countercyclical capital buffers (Drehmann et al., 2010). Despite its importance, it remains a challenge to estimate the output gap. Since it is unobservable by nature, the results depend on the assumptions of the underlying model. Furthermore, many methods produce estimates that are subject to significant fluctuations at the end of the sample due to data revisions (Orphanides and van Norden, 2002; Orphanides, 2003b). This is a problem for policymakers that rely on stable and economically meaningful real-time estimates. Our article contributes to the literature in various ways. First, we extend the error decomposition framework in Jacobs et al. (2022) to the case where multiple output gap estimates are reconciled, thus providing a 'true' output gap. Second, we treat the output of various models as noisy estimates of the unobserved cyclical position, which allows us to decompose the revisions for each underlying method into noise and news shocks. Using the Kalman gain, we are also able to evaluate the information content each method contributes to the reconciled gap. Third, we provide a comprehensive real-time evaluation of our proposed reconciliation model, showing that our measure is resilient to revisions in real-time compared to competing approaches.

The third Chapter examines the role of communication shocks originating in press releases of the ECB. Central bank communication plays an important role in the conduct of modern monetary policy (Coenen et al., 2017; Swanson, 2017). We use a unique data set that encompasses almost 43,000 classified statements which stem from the introductory statements after each ECB Governing Council meeting. We use the statements to build three indicators related to the monetary policy mandate of the ECB, namely price stability, the real economy, and monetary phenomena. Our results indicate a strong and robust impact of communication regarding the real economy and price stability on the shadow rate, which reflects the assessment of financial market participants of the overall monetary policy stance (Krippner, 2013a,b). We also contribute to the literature by showing that professional forecasters react to the communication of the ECB, despite controlling for various variables related to the (expected) state of the economy and the actual interest rate change.

The fourth and last Chapter of this dissertation deals with one of the largest shocks in recent times, namely the COVID-19 pandemic. Using administrative data from the phone network provider, me and my co-author study the mental health effects in Switzerland as proxied by calls to the helpline "Offering a Helping Hand". The advantage of helpline data is its availability at high frequency and the fact that it may serve as a measure of mental health in the general population. This is true especially in times of lockdowns when establishing physical contact, for example with a psychologist, is more complicated if not impossible. Thus, the apparent suitability of call data has spurred research in this area after the outbreak of the pandemic (Monreal-Bartolomé et al., 2022; Batchelor et al., 2021; Brülhart et al., 2021; Zalsman et al., 2021; Turkington et al., 2020). We contribute to the literature in two important dimensions. First, we do not rely on manually recorded data entries, which often (as we show in the case of Switzerland) features a biased and rather incomplete sample of calls. Instead, we use automatically recorded administrative data on the phone level, which is a comprehensive data source and also allows us to track individual callers. This gives us the opportunity to identify unmet need for psychological counsel, which is a result of capacity constraints as we find. Second, while the first COVID-19 wave led to a significant increase in call volumes and duration to the helpline, we do not find such a response during the second wave. We explain this by the fact that social distancing policies were much less restrictive during the second wave.

Chapter 1

Housing Demand Shocks and Households' Balance Sheets

1.1 Introduction

The financial crisis of 2007-2008 was closely related to the collapse of the U.S. housing market. Accordingly, a growing literature suggests that housing demand shocks are significant drivers of business cycle fluctuations (Iacoviello and Neri, 2010; Liu et al., 2013).

The tight connection between economic activity and housing is often ascribed to three main channels. The first one stems from Tobin's q effect, which indicates that rising house prices lead to an increase in real estate values over construction costs, thereby stimulating residential investment which can be a key component of GDP growth (Leamer, 2007). A second transmission channel works via the credit market. A positive housing demand shock increases homeowners' assets and thus the value of collateral, mitigating households' and firms' credit constraints (Iacoviello, 2005; Mian and Sufi, 2011; Liu et al., 2013). The same balance sheet mechanism applies to the lending side, as financial intermediaries expand credit supply along with the increase of assets held (Adrian et al., 2010). The third channel is an implication of consumption smoothing in life-cycle models. An unexpected and permanent increase in wealth leads households to borrow more and increase consumption. The dynamic interdependence between household balance sheets, the credit market and the housing market is thus able to amplify each channel, possibly making it a powerful driver of real economic activity.

The experience of the Great Recession indicated that the strength of such feedback loops also depends on the heterogeneity of households' balance sheets. While traditional macroeconomic models assume a representative agent framework in which households are insured against individual wealth shocks, recent empirical evidence points at the importance of modelling housing, credit and consumption using heterogeneous agent models (Bostic et al., 2009; Mian and Sufi, 2011; Aladangady, 2017). Of particular interest has been the relationship between a shock to housing wealth and the cross-sectional response of consumption, as housing market spillovers are suggested to be largely concentrated on consumption (Iacoviello and Neri, 2010, see also channel 2 and 3 above). A prominent example is Mian et al. (2013), who show that an increase of housing wealth implies markedly different consumption responses depending on households' initial net wealth. Berger et al. (2017) develop a quantitative model that stresses the importance of individual households' borrowing constraints in explaining the large reaction of aggregate consumption.

A proper understanding of the joint dynamics of the real economy, the financial sector and households' balance sheets is thus crucial in forming more preemptive policy measures, be it fiscal or monetary. Given the many variables involved, a comprehensive empirical analysis requires an econometric framework that is able to sufficiently model such rich joint dynamics.
This article reexamines the empirical evidence concerning housing demand shocks and its propagation mechanisms in a dynamic environment consisting of a broad range of macroeconomic and financial series. To show robustness of the results, we use a smaller dataset consisting of 73 variables and a larger one with 160 series.

Both datasets comprise various balance sheet components for households belonging to four different wealth percentile groups. We estimate dynamic factor models (DFMs), which allow to extract a few latent common factors from a large panel of economic series, thus capturing the underlying information without running into degrees of freedom problems. This is especially important in the case considered here, as the impact of a housing shock on household balance sheet components is less likely to be homogeneous across wealth groups. The dynamic effects of a housing demand shock are then obtained by imposing a set of theoretically motivated sign restrictions on the contemporaneous response of a few selected economic indicators and a large number of housing variables.

The paper's main contributions to the literature are the following. First, the identification of the housing demand shock is arguably sharper than in the existing literature. The vast majority of studies impose potentially strong identifying assumptions by employing a temporal ordering to the impact matrix while using quarterly data (Jarocinski and Smets, 2008; Cardarelli et al., 2009; Bagliano and Morana, 2012; Cesa-Bianchi, 2013; Buch et al., 2014). Based on the transmission channels described above, however, it seems likely that all economic indicators react within one quarter when the housing market is disrupted. While Furlanetto et al. (2019) identify a housing shock using only sign restrictions, it is not possible to analyze the effects on the housing market due to the use of a six-variable vector autoregression (VAR). Our identification scheme relaxes the possibly contaminated zero restrictions, while the model environment allows to restrict and analyze the response of a broad set of housing series. Second, for four different wealth percentile groups, we assess the reactions of balance sheet components when faced with a housing demand shock. This allows to analyze not only if the dynamic consumption and credit response differs with respect to the initial level of net wealth, but also whether the distribution of net wealth gains is uniformly distributed across wealth segments. A historical decomposition enables to crosscheck the popular narrative that negative housing demand shocks have caused poorer households to cut consumption sharply during the Great Recession, while households at the top of the wealth distribution were unaffected. Finally, we compute the elasticity of consumption with respect to house prices for each wealth group and compare it to the literature on wealth effects. To the best of our knowledge, this is the first study to analyze these effects in a widely data-driven and dynamic environment. It is thus highly complementary to the empirical literature on the heterogeneous effects of housing shocks using microdata (Bostic et al., 2009; Mian and Sufi, 2011; Mian et al., 2013; Aladangady, 2017), which often relies on cross-regional regressions to obtain partial equilibrium estimates of the consumption or credit response. While we are concerned with the general equilibrium effects of a housing demand shock, our DFM imposes little structure a priori in contrast to studies on the wealth effect of housing using DSGE models (for example, Berger et al., 2017; Guren et al., 2021).

Our results show that an unexpected positive housing demand shock causes a large and persistent boom of real economic activity and a brightening of labor market conditions, despite an immediate increase of interest rates. The expansion of credit and consumption aggregates along with house prices suggests that both the credit mechanism and the wealth effect are important transmission channels of housing shocks. The strong positive response of all unrestricted housing start series and residential investment provide some comfort to the identifying assumptions. The second part of our analysis provides evidence that housing demand shocks are key to understand household balance sheet dynamics across the wealth distribution. In particular, impulse responses suggest that the reaction of durable consumption is a decreasing function of net wealth. Variance decompositions assign a bulk of the variation found in consumption to the housing shock, especially for the bottom 50% of households. The historical decomposition shows that the cumulative effects of housing demand shocks have significantly affected consumption of the poorest during the Great Recession. However, we also find evidence that part of the cutback in top 1% consumption is explained by housing demand shocks in that period. Elasticities of consumption with respect to house prices, which are a byproduct of our structural model, suggest a larger spending sensitivity for weaker balance sheet households. Our results are highly robust across both datasets.

The rest of the paper is organized as follows. Section 2 discusses the econometric framework and structural identification of the housing demand shock. The data are outlined in Section 3. Impulse response functions and variance decompositions are analyzed in Section 4. Section 5 concludes. The Appendix provides additional results with alternative model specifications and gives a detailed overview of the data.

1.2 Econometric Framework

Structural macroeconomic shocks are often identified using small-scale VAR models in order to conserve degrees of freedom. The inclusion of only a couple of series, however, is likely to cause omitted variable problems that may affect impulse responses and variance decompositions alike (see Bernanke et al., 2005 for a thorough discussion). Moreover, economic concepts such as "real activity" are hard to capture using only a specific data series. This issue is especially pronounced in the case considered here. Using only one or two housing series to identify a housing shock introduces an arguably large degree of arbitrariness. Finally, VARs allow to compute impulse responses only for the included variables, which is often a subset of what the researcher or policy-maker is interested about.

To take advantage of the information contained in large panel datasets we employ a dynamic factor model (see Stock and Watson, 2016 for an overview). Intuitively, the DFM approach boils down to extracting a small number of latent common factors that summarize the information from a much richer dataset. Consider the following dynamic factor model

$$y_t = \sum_{l=0}^{s} \lambda_l f_{t-l} + e_t \qquad \text{with } e_t \sim iid.N(0, R) \tag{1.1a}$$

$$f_t = \sum_{p=1}^{h} \phi_p f_{t-p} + \epsilon_t \qquad \text{with } \epsilon_t \sim iid.N(0, Q) \tag{1.1b}$$

where y_t is a $(k \times 1)$ vector of observed variables and f_t is a $(m \times 1)$ vector of latent factors. The number of factors is typically much smaller than the length of the cross-section, $m << k$. The idiosyncratic disturbances e_t are uncorrelated at all leads and lags with f_t and ϵ_t and are assumed to have a diagonal covariance matrix R. The $(k \times m)$ loading matrices λ_l relate the factors to all economic and financial indicators. Note that if $s > 0$ then the factor lags are allowed to directly affect the series in y_t. Finally, ϕ_p are coefficient matrices, each of dimension $(m \times m)$, governing the dynamics in the state equation (1.1b). Clearly, the state equation follows the common VAR structure. The factor model can also be expressed in static form (see Stock and Watson, 2016) if we let $h \geq s+1$, $F_t = (f_t', \ldots, f_{t-h+1}')'$ and $\Lambda^* = (\lambda_0, \ldots, \lambda_s)$:

$$y_t = \Lambda F_t + e_t \tag{1.2a}$$

$$F_t = \Phi F_{t-1} + u_t \tag{1.2b}$$

where $\Lambda = \begin{pmatrix} \Lambda^* & 0_{k \times m(h-s-1)} \end{pmatrix}$, $\Phi = \begin{pmatrix} \phi_1 & \cdots & \phi_h \\ I_{m(h-1)} & & 0_{m(h-1) \times m} \end{pmatrix}$ and $u_t = \begin{pmatrix} \epsilon_t \\ 0_{m(h-1) \times 1} \end{pmatrix}$.

Note that in contrast to studies using factor-augmented VAR (FAVAR) models, we do not impose that any factor be observed (for example, Bernanke et al. (2005) impose a short-term interest rate as observed factor and use this assumption for identificaton of the structural shock). For our purposes, it is also not necessary to provide an economic interpretation to the latent factors as in Belviso and Milani (2006).

1.2.1 Estimation

We rely on now standard Bayesian methods to estimate the model in (1.2a-1.2b). The likelihood-based approach is able to explicitly exploit the structure of the state equation in the estimation of the factors (see Bernanke et al., 2005). As a byproduct, it also delivers full posterior distributions naturally quantifying the extent of uncertainty. The estimation procedure consists of a multimove Gibbs sampler (Carter and Kohn, 1994, see Kim and Nelson, 1999 for an overview) in which the parameters are sampled conditional on the most recent draws of the factors, and the factors are drawn conditional on the most recent parameter estimates. Following Bai and Wang (2015), we employ the Jeffreys prior to account for the lack of a priori information about the model parameters. The Bayesian framework considered here is thus able to perform estimation and inference of large dynamic factor models with a dynamic structure in both the observation and the state equation.

It is well known that due to indeterminacy the factor model is not identified without further restrictions (see Stock and Watson, 2016). The unobserved factors can always be rotated into an observationally equivalent representation. To achieve identification, we rely on the minimal restriction conditions derived in Bai and Wang (2015). More exactly, we assume that the upper $(m \times m)$ block of λ_0 is an identity matrix. Since the identifying restrictions concern only the contemporary loadings, the VAR(h) dynamics of the factors are left completely unrestricted. In other words, it is not necessary to impose the conventional practice that the covariance matrix of the dynamic factors f_t is an identity matrix. A disturbance arising from the state equation is thus able to affect all macroeconomic and financial indicators in y_t with no lags. The use of this minimal restriction strategy allows to impose contemporaneous sign restrictions on any series to identify our structural housing demand shock.

1.2.2 Structural Identification

We assume that the reduced-form vector of state-equation shocks ϵ_t is a linear combination of structural shocks w_t, such that $\epsilon_t = A_0 w_t$. Under the assumption that $\mathbb{E}[w_t w_t'] = I_m$, we can decompose the covariance matrix $Q = \mathbb{E}[\epsilon_t \epsilon_t'] = \mathbb{E}[A_0 w_t w_t' A_0'] = A_0 A_0'$. One popular approach to fully identify the system is to choose A_0 to be the Cholesky factor of the VAR covariance matrix as in Sims (1980). It is well known that the triangular nature of the Cholesky factor implicitly imposes a temporal ordering (see, for example, the discussion in Canova and De Nicolo, 2002). In the context of housing shocks, this identification scheme has been followed by Cesa-Bianchi (2013) and Bagliano and Morana (2012). In contrast, Buch et al.

(2014), Cardarelli et al. (2009) and Jarocinski and Smets (2008) impose a mixture of sign and zero restrictions, assuming that output or prices react to housing shocks only with a lag. Given the many channels through which the housing market may affect the economy, such zero restrictions are arguably imposing strong identifying assumptions. For that reason, we adopt the agnostic identification strategy by Uhlig (2005) and, for the factor model framework, Amir-Ahmadi and Uhlig (2015), and only restrict the sign of selected variables.

All sign restrictions are theoretically motivated and are keeping with a large literature summarized in Justiniano et al. (2015), Christiano et al. (2014) and Furlanetto et al. (2019). More specifically, we identify the housing demand shock by imposing that real GDP, core prices of personal consumption expenditures (PCE) and the three month treasury bill rate react positively upon impact. This rules out that the identified shock is contaminated with supply or monetary shocks. It is, however, more challenging to disentangle the housing demand shock from investment shocks (see Justiniano et al., 2010) and other demand shocks (for example discount factor or fiscal shocks). Our strategy follows Furlanetto et al. (2019), who suggest that traditional demand shocks can be excluded by imposing a positive sign restriction on the response of the ratio of (real gross private domestic) investment over output.[1] Investment and housing shocks are thus assumed to create investment booms. To further separate housing from investment shocks we restrict stock prices as measured by the S&P 500 to react positively.[2] Note that up to now, it is still not possible to distinguish a positive housing shock from a shock originating in the credit markets. Keeping with Furlanetto et al. (2019) and the literature in Justiniano et al. (2015), we restrict the loan-to-value (LTV) ratio to react negatively. This allows total credit to increase under the restriction that the value of real estate does so by

[1]Restricting the ratio allows investment to increase when faced with a demand shock but less so than the remaining part of aggregate demand.

[2]A positive investment shock increases the supply of capital and lowers its price. Since the price of capital is taken as a proxy of the stock market value for firms, an investment shock negatively impacts the value of equity (see the discussion in Furlanetto et al., 2019 and Christiano et al., 2014).

more.[3] Finally, a positive housing demand shock increases real estate loans as well as quantity and prices of housing (Iacoviello and Neri, 2010). We therefore use our detailed dataset to impose positive sign restrictions on aggregate real estate loans as well as on building permits and house prices in all four Census regions.[4] Since a shock is unexpected by nature, it is arguably more appropriate to constrain permits instead of housing starts, which are left unrestricted. In contrast to small-scale VAR studies also using sign restrictions (André et al., 2012, Furlanetto et al., 2019), our factor model allows testing whether the identified housing shock actually leads to new housing units being built.

To incorporate the sign restrictions we employ the procedure described in Rubio-Ramirez et al. (2010). For each draw from the reduced-form posterior, the algorithm works as follows. First, we produce A_0^{chol} using the Cholesky decomposition of the covariance matrix Q to obtain uncorrelated shocks as in a recursively identified model. Then we compute the QR decomposition of $W = QR$, where W is a $(m \times m)$ matrix filled with independent standard normally distributed variables. Post-multiplying A_0^{chol} by Q forms a candidate structural impact matrix $A_0^{cand} = A_0^{chol}Q$. Contemporaneous impulse responses are then generated according to $y_0^{irf} = \Lambda^* A_0^{cand}$, which allows to check the sign restrictions. If this is not the case we draw a new W and repeat the procedure until the sign restrictions are satisfied. Note that we follow the recommendations of Canova and Paustian (2011) in two ways. First, we restrict only impact responses to avoid the often questionable dynamic constraints. Second, the factor model framework allows us to impose a large number of sign restrictions to identify the structural shock. Moreover, all sign restrictions are motivated by the dynamic stochastic general equilibrium (DSGE) literature as advised by Musso et al. (2011).

[3]More exactly, total credit amounts to total loans to the non-financial private sector, while housing value is given by real estate at market value of households and non-profit organisations.

[4]The US Census Bureau splits the U.S. into four regions: the Northeast, the Midwest, the South, and the West.

1.3 Data

We estimate the dynamic factor model on two quarterly datasets, one containing 73 and the other 160 indicators. As suggested by Bai and Ng (2008a) and Boivin and Ng (2006), a larger cross-section may not necessarily improve estimation accuracy in factor models. It thus seems appropriate to use both a relatively small dataset consisting of only key series and a large one to check the robustness of the results. The sample starts in 1989Q3 and ends in 2019Q3. All indicators are transformed to induce stationarity, see the Appendix for a detailed description. The large dataset is an augmented version of the FRED-QD database (see McCracken and Ng, 2020) with the main difference that household balance sheet data has been added. The smaller dataset builds on the same database but focuses on key economic indicators, the housing and credit market and household balance sheet data.

To evaluate whether a housing demand shock has heterogeneous effects on households' balance sheets we draw on the Distributional Financial Accounts (DFAs) provided by the Board of Governors (see Batty et al., 2019). The relatively new DFAs contain detailed time series on the assets and liabilities held by four different wealth percentile groups (bottom 50%, 50th to 90th percentiles, 90th to 99th percentiles, and the top 1%). Figure 1.1 shows the evolution of real net wealth for all wealth percentile groups. The greatest wealth losses are found between 2007Q3 and 2009Q1 and are especially concentrated on households belonging to the top decile. The subsequent recovery is associated with an unprecedented wealth level, however unequally distributed among the population. While net worth of the bottom 50% is close to zero for the entire sample, the top 1% accounts for a substantial fraction of U.S. wealth. In 2019Q3, around 70% of wealth is held by the top decile. The heterogeneous evolution among wealth groups is clearly related to individual composition effects. Wealthier households have traditionally been exposed more to cyclical assets like equity, as the largest fraction of wealth is of financial nature (see Figure 1.2 for 2019Q3). Real estate is the principal asset component for households belonging to the bottom 50%. In contrast to other percentile groups, financial assets account for only a minor share of total assets. The liability side shows the poorest households'

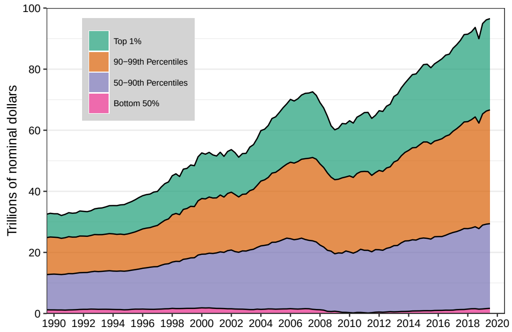


Figure 1.1: Evolution of Net Worth

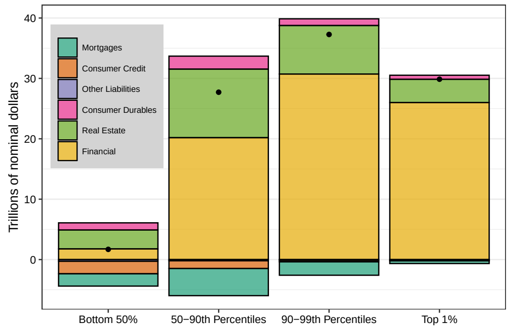


Figure 1.2: Balance Sheet Composition 2019Q3

dependence on consumer credit, as they hold over half of total consumer credit.

In order to ensure that results are not driven by population differences from the different-sized wealth groups, we scale the series of each group by the respective number of percentiles.[5] Moreover, all series enter the estimation procedure in standardized form.

1.4 Results

In this section, we present empirical results from our dynamic factor model with both datasets. To determine the number of unobserved factors, m, we compute the information criteria IC_{p2} and BIC_3 by Bai and Ng (2002) (see also the discussion in Bai and Ng, 2008b), the ER and GR criteria by Ahn and Horenstein (2013) and the $IC^{T*}_{1,c,n}$ and $IC^{T*}_{2,c,n}$ criteria by Alessi et al. (2010). For the smaller dataset, all criteria suggest five factors except for $ER = 2$ and $IC_{p2} = 8$. The same two criteria are estimated to be $ER = 2$ and $IC_{p2} = 6$ with the rich dataset, while all other suggest four factors. We follow the majority and set $m = 5$ for the narrow and $m = 4$ for the more extensive database, but report results for a range of alternative specifications in the Appendix. We set $s = 1$ in observation equation (1.1a), thus

[5]For example, credit held by the bottom 50% is divided by 50.

allowing a dynamic relationship between the latent states and the variables. According to the Bayesian information criterion (BIC), the factor dynamics in (1.1b) are characterized by a VAR(1) for the smaller and a VAR(2) for the larger dataset. We follow Uhlig (2005) and report the pointwise median together with 68% posterior bands. The y_t vector has been ordered such that the first m series are given by leading indicators of different data groups.[6] All results are based on 90'000 draws from the Gibbs sampler, from which the first 87'000 are discarded as burn-in.

The presentation of our empirical results is organized around our two research questions. We begin with the question of how the housing demand shock is transmitted to the aggregate US economy before analyzing the responses of households' balance sheets.

1.4.1 How Are Housing Demand Shocks Transmitted to the Aggregate Economy?

Impulse Responses

Figure 1.3 plots the impulse responses, estimated on the smaller sample, of the level of key indicators to a one standard deviation housing demand shock. The black line reports the median response. Results from the large dataset are vastly similar and therefore relegated to the Appendix (see Figure 1.8). In line with the transmission channels discussed above, a housing demand shock generates a significant and persistent economic boom. Real GDP, industrial production, and capacity utilization increase between 0.5 and 0.6% within one year. Residential investment, a subcomponent of GDP measuring the purchase of newly built residential structures, rises disproportionally by 2% after four quarters. The economic expansion is reflected in the labour market indicators, as the unemployment rate decreases while the total

[6] More specifically, the first five series consist of real GDP, the industrial production index, the total number of nonfarm employees, total housing starts, and the S&P 500. See the Appendix for a detailed list of the groups and series. Random permutations of the observation vector had negligible effects on our results.

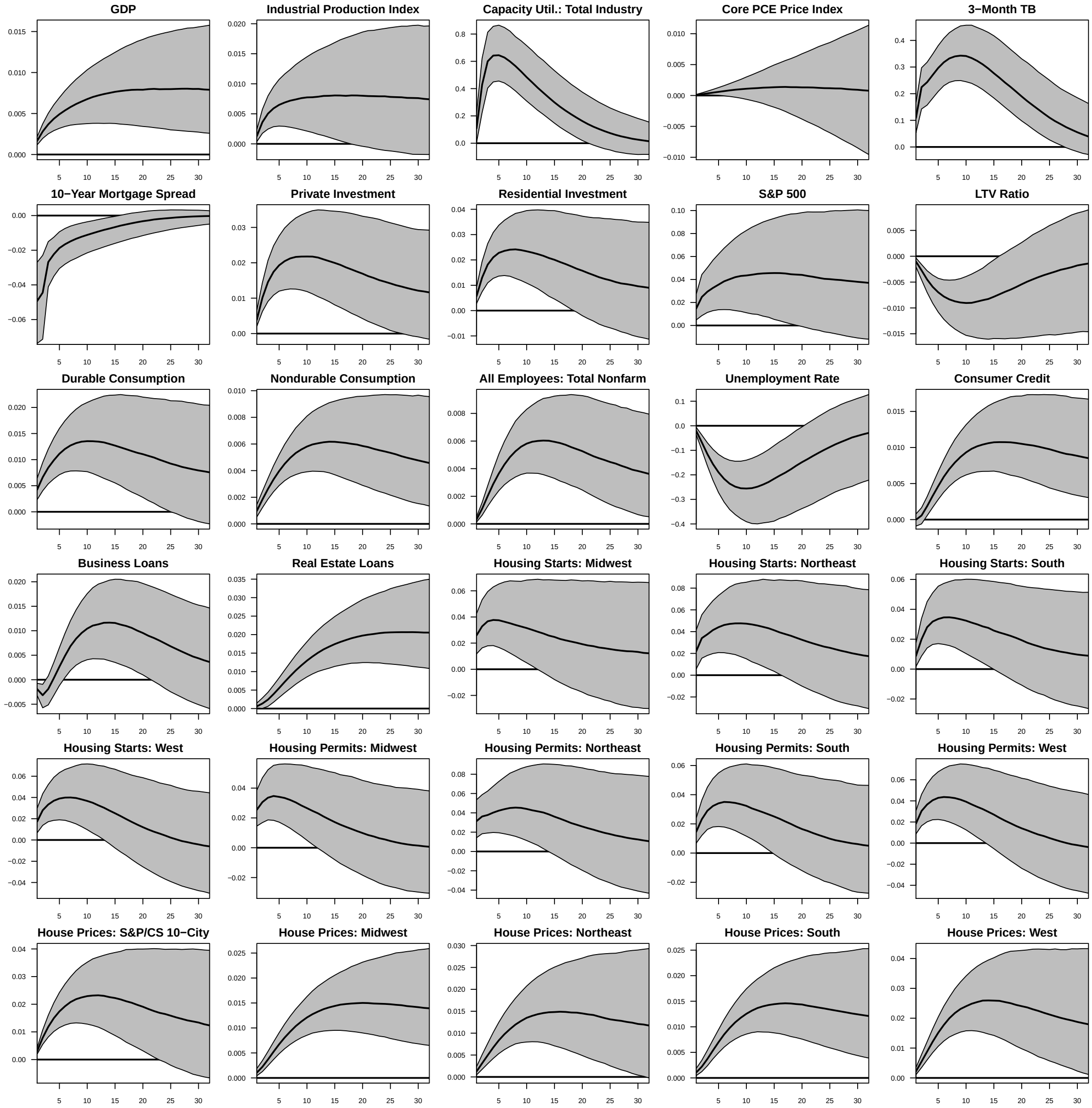


Figure 1.3: Dynamic responses of variables to housing demand shock using the small dataset. Notes: The figure plots the impulse responses of the level of key variables to the identified housing demand shock. The gray areas indicate the 68% posterior probability regions. The straight black line indicates the posterior median at each horizon.

number of nonfarm employees grows. On the credit side, real estate loans and consumer credit increase persistently over the impulse horizon, which conforms to the collateral channel of a housing demand shock. The fact that business loans grow only slightly and return to their initial level stands in contrast to studies identifying credit shocks (for example, Boivin et al., 2018), thus pointing towards a successful disentanglement of credit and housing shocks. The LTV ratio is restricted to be negative on impact but stays so for at least four years. Conforming to consumption smoothing arguments from life-cycle models, both durable and nondurable consumption rise. Importantly, the response of durable consumption is more pronounced in magnitude than for nondurable consumption. Financial variables mirror the boom of real economic indicators. The 3-month treasury bill rate is hump-shaped, suggesting that the Fed increases its benchmark rate as a reaction to the initial rise of GDP growth and the capacity utilization index. The median response of the S&P 500 stock market index increases by around 4% over the horizon. Interestingly, the reaction of the consumption price index is relatively muted despite the economic expansion, confirming the empirical results in Furlanetto et al. (2019). More exactly, the core PCE price index is slightly positive for only a few quarters, after which the posterior distribution centers around the zero line. In contrast, house price indices of all four Census regions react strongly positive to the housing demand shock over the entire horizon. The S&P/Case-Shiller 10-City home price index, which was left unrestricted also on impact, confirms the price hike. In line with Tobin's q effect, rising prices for housing are associated with a general increase of building permits for at least three years. The fact that the unrestricted housing starts rise in all four geographic areas provides further comfort to our identifying assumptions.

All impulse responses discussed are robust to the choice of the dataset, even though the model specification differs and the extensive dataset contains more than twice as much variables as the smaller one.[7] Among these additional series are sectoral employment indicators, which provide further evidence that a housing shock has in fact been identified. Figure 1.9 in the Appendix plots the respective impulse responses and shows that the construction sector features the most pronounced reaction among all sectors.

[7]Further robustness tests are reported in Appendix Figures 1.11 and 1.12, covering a wide range of alternative model specifications.

Variance Decompositions

Table 1.1 shows the importance of housing demand shocks in explaining economic fluctuations during our 1989-2019 sample. While variance decompositions are calculated for each draw that satisfies the sign restrictions, the table presents median values at a 32-quarter horizon. The second column reports the fraction of the common component of the series explained by the housing demand shock. The third column shows the contribution of the common component to the overall variance of the indicators. The fourth column is simply the product of the previous two and represents the fraction of the total variability explained by the housing demand shock. The next three columns show the results for the large dataset. In line with the literature, our variance decompositions show that housing demand shocks are significant drivers of business cycle fluctuations. The housing shock explains between 24 and 44% of the variance of the common component of real economic indicators such as GDP, the industrial production index and the capacity utilization measure. Labor market indicators, credit series and the 3-month TB rate are also well explained. Importantly, the housing demand shock has high explanatory power for housing starts, housing permits and house price variation for all four Census regions. Column four and seven indicate that our structural model explains around 20% of total variation in house price series. Column three and six of Table 1.1 show that five and four unobserved factors are able to capture a large fraction of fluctuations in key economic indicators.

1.4.2 Do Households Respond Differently to the Housing Demand Shock Depending on Net Wealth?

The substantial aggregate effects of a housing demand shock demonstrate the importance of understanding the underlying transmission channels. Of particular interest after the financial crisis has been the relationship between fluctuations in housing wealth and the response of consumption. In fact, housing market spillovers are suggested to be largely concentrated on consumption (Iacoviello and Neri, 2010). As outlined above, an increase of housing wealth may a) raise collateral and therefore

Series	Small Dataset VD	Small Dataset R^2	Small Dataset Total	Large Dataset VD	Large Dataset R^2	Large Dataset Total
GDP	0.33	0.66	0.22	0.43	0.68	0.29
Industrial Production Index	0.13	0.74	0.10	0.33	0.82	0.27
Capacity Util.: Total Industry	0.36	0.93	0.33	0.44	0.91	0.40
Core PCE Price Index	0.09	0.19	0.02	0.11	0.27	0.03
3-Month TB	0.47	0.97	0.45	0.56	0.94	0.53
10-Year Mortgage Spread	0.22	0.53	0.11	0.26	0.52	0.14
Private Investment	0.18	0.64	0.11	0.26	0.64	0.17
Residential Investment	0.18	0.75	0.13	0.36	0.63	0.22
S&P 500	0.16	0.72	0.12	0.25	0.45	0.11
LTV Ratio	0.17	0.70	0.12	0.27	0.61	0.17
Durable Consumption	0.20	0.46	0.09	0.37	0.46	0.17
Nondurable Consumption	0.30	0.42	0.12	0.41	0.47	0.19
All Employees: Total Nonfarm	0.23	0.94	0.22	0.26	0.93	0.25
Unemployment Rate	0.20	0.82	0.16	-	-	-
Unemployment Rate: Women	-	-	-	0.30	0.70	0.21
Consumer Credit	0.27	0.41	0.11	0.42	0.37	0.16
Business Loans	0.14	0.62	0.09	0.23	0.53	0.12
Real Estate Loans	0.23	0.52	0.12	0.39	0.41	0.16
Housing Starts: Midwest	0.17	0.32	0.05	0.29	0.22	0.06
Housing Starts: Northeast	0.10	0.46	0.05	0.21	0.35	0.07
Housing Starts: South	0.13	0.50	0.07	0.31	0.43	0.13
Housing Starts: West	0.16	0.46	0.07	0.35	0.34	0.12
Housing Permits: Midwest	0.16	0.70	0.11	0.35	0.50	0.17
Housing Permits: Northeast	0.13	0.40	0.05	0.26	0.27	0.07
Housing Permits: South	0.14	0.74	0.10	0.33	0.54	0.18
Housing Permits: West	0.17	0.52	0.09	0.34	0.39	0.13
House Prices: S&P/CS 10-City	0.23	0.81	0.19	0.36	0.74	0.27
House Prices: Midwest	0.29	0.67	0.20	0.43	0.60	0.26
House Prices: Northeast	0.18	0.75	0.13	0.28	0.66	0.19
House Prices: South	0.26	0.77	0.20	0.38	0.65	0.25
House Prices: West	0.26	0.85	0.22	0.37	0.71	0.26

Table 1.1. Variance Decompositions and R^2. Notes: The second column reports the fraction of the common component of the series explained by the housing demand shock at a 32-quarter horizon. The third column shows the contribution of the common component to the overall variance of the indicators. The fourth column is simply the product of the previous two and represents the fraction of the total variability explained by the housing demand shock. The next three columns show the results for the large dataset.

mitigate credit constraints and b) increase consumption due to the wealth effect. Both channels are arguably sensitive to the cross-sectional distribution of wealth and the composition of households' balance sheets. For example, richer households are less likely to face credit constraints and may give the precautionary savings motive a small role. This translates into a smaller elasticity to consume out of wealth compared to the poor. As a consequence, the aggregate consequences of a housing shock might largely be driven by the consumption response of poorer households, although housing is a substantial balance sheet component for all wealth percentile groups.

Given the economic importance, an extensive number of studies have analyzed whether and to what extent (housing) wealth shocks cause heterogeneous responses among households. The hypothesis of a representative agent is typically rejected and economically significant wealth effects are reported (see, for example, Mian and Sufi, 2011; Mian et al., 2013; Aladangady, 2017; Berger et al., 2017; Guren et al., 2021). To isolate the housing wealth effect, empirical studies frequently rely on cross-regional regressions using exogeneous sources of housing variation as instruments. This strategy delivers partial equilibrium estimates, as aggregate general equilibrium effects are absorbed by the constant or the time fixed effect in a panel regression.[8] In contrast and highly complementary to this strand of literature, we are interested in the general equilibrium effects of housing shocks. Looking again at the consumption response of different wealth percentile groups, it is interesting to see whether the implications of partial equilibrium estimates still hold. For example, it is well possible that the financial wealth effect associated with booming stock markets ignites consumption of the rich, thus leveling the cross-sectional consumption response. To understand the aggregate general equilibrium dynamics in the last section we thus analyze the general equilibrium behavior of households' balance sheets. In contrast to studies using DSGE models (Berger et al., 2017, Guren et al., 2021, among others), our DFM imposes little structure a priori, thus mitigating the possibility that results are driven by a potentially misspecified model structure. Clearly, this comes at the cost of not being able to obtain "deep" structural parameters, for example the wealth effect of housing values conditional on household characteristics. It is also important to note a drawback concerning the disaggregated consumption measures used in this study. The use of household balance sheet data implies that nondurable consumption is not available, so that only the reaction of durable consumption can

[8]As Guren et al. (2020) point out, the clear theoretical interpretation of cross-regional partial equilibrium estimates might be contaminated by local general equilibrium effects. They suggest to divide cross-regional estimates by an estimate of the local fiscal multiplier to adjust for local general equilibrium effects.

be estimated. However, comparing both aggregate consumption responses in the last section shows a relatively muted reaction for nondurable consumption in terms of magnitude. Durable consumption and (residential) investment are the primary forces driving the pronounced GDP response. The literature also suggests that increased spending due to housing wealth is largely concentrated on durable goods (Mian et al., 2013; Andersen and Leth-Petersen, 2019).

Impulse Responses

Estimated on the small sample, Figure 1.4 plots impulse responses of the level of real household balance sheet components to the housing demand shock. The first five columns show assets and liabilities, while the last column plots the response of total net wealth. Rows list the four wealth percentile groups. Results from the large dataset are almost identical,[9] see Appendix Figure 1.10. Note that responses are highly data-driven, as all balance sheet series are left completely unrestricted. A quick glance is enough to observe that a housing demand shock is associated with a general expansion of balance sheets across all wealth segments. The increase of assets generally dominates the higher debt burden, although net wealth for households belonging to the bottom 50% slowly revert to the initial state. All groups participate in the newly built structures, as the real estate value held by each segment rises somewhat more than house price indices in the last section. However, the median response is most pronounced for households in the bottom 50% with a final growth of around 3.5%. In line with the stock market boom, financial assets held by all wealth segments increase. For both groups in the top decile, the response of financial assets clearly drives total net wealth dynamics (in 2019Q3, 85% and 77% of total assets are of financial nature for the top 1% and the 90-99th percentile group, respectively). On the credit side, we would expect a stronger tendency to borrow among poorer households, as net wealth is likely to proxy for credit constraints. In fact, Mian and Sufi (2011) provide evidence that only households at the top of the credit score distribution do not increase total debt leverage when faced with

[9]Most impulse responses are slightly more pronounced with the large dataset. Appendix Figures 1.13 and 1.14 confirm the benchmark results for a wide range of alternative model specifications.

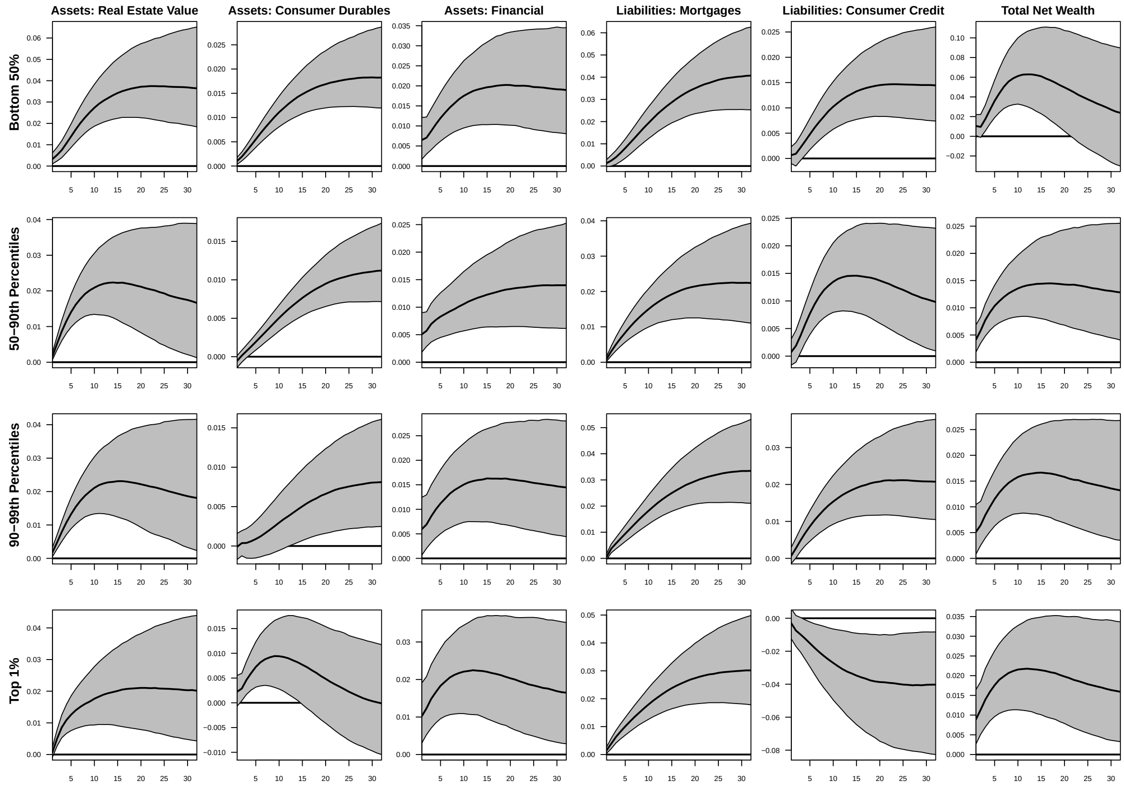


Figure 1.4: Dynamic responses of household balance sheet variables to housing demand shock using the small dataset. Notes: The second column reports the fraction of the common component of the series explained by the housing demand shock at a 32-quarter horizon. The third column shows the contribution of the common component to the overall variance of the indicators. The fourth column is simply the product of the previous two and represents the fraction of the total variability explained by the housing demand shock. The next three columns show the results for the large dataset.

a housing wealth shock. Our impulse responses imply a significant and persistent expansion of mortgages across the wealth spectrum. Heterogeneity is not especially accentuated, although median mortgage expansion is largest for the bottom 50%. Comparing the evolution of mortgages and credit, we find equity extracted from housing to be more pronounced than non-housing related consumer credit for all groups (note that home equity lines of credit are also included in the mortgage category). Our findings support the evidence that home equity-based borrowing is not used to pay down expensive consumer credit balances (Mian and Sufi, 2011), as only the top 1% chooses to deleverage.

The real consequences of even a broad credit growth, however, ultimately depend on how households use their new liquidity. If most households just keep new money on the bank account, we would arguably observe little change in real outcomes. Apart from residential construction, the reaction of consumption is key to understand the strong linkage of housing shocks and business cycle fluctuations. The obtained impulse responses for durable consumption suggest that household balance sheet strength plays an important role in determining aggregate demand. The estimated long-term increase for the bottom 50% of households centers around 1.75%, while the effect on the next 40% is in the range of 1%. Households in the 90th to 99th percentiles increase durable consumer spending by about 0.75%, although the posterior mass concentrates above the zero line only 3 years after the shock. For the richest percentile of households, the housing demand shock first raises durable consumption before reverting back to the initial value. Conditional on the stock market boom and the remarkable holdings of financial assets of the top 1%, the muted reaction of consumption suggests a small role for financial wealth in determining consumption fluctuations. This conforms to the literature estimating financial wealth effects, which are reported to be substantially smaller than housing wealth effects (Carroll et al., 2011; Bostic et al., 2009; Slacalek, 2009; Muellbauer, 2008). In summary, the response of aggregate durable consumption to a housing demand shock is estimated to be a decreasing function of net wealth. The evidence provided by the literature that spending of the rich does not react much to a housing wealth shock thus carries over to the factor general equilibrium case. However, given the notable net wealth of 50th to 90th wealth percentile households, one may wonder why the respective response of durable consumption is pronounced in the first place. Following Kaplan et al. (2014), a possible explanation is that a fraction of households belonging to this wealth segment are "wealthy hand-to-mouth". Despite owning sizable amounts of assets, these households hold balance sheets that carry little liquid wealth. As a result, they are more susceptible to wealth shocks than non hand-to-mouth households.[10]

[10]Households in the 50th to 90th wealth percentiles held 64% of total assets in either real estate or pension funds in 2019Q3, which is the highest share among all groups (Batty et al., 2019).

Variance Decompositions

Table 1.2 shows the importance of the housing demand shock in explaining the variance of balance sheet components for each wealth group. Irrespective of the dataset, the housing shock is a significant driver of the common component of all balance sheet indicators. The variance explained of total net wealth is sizeable, depending on wealth group and dataset between 21 and 42%. The housing demand shock not only has high explanatory power for fluctuations of real estate and mortgage components, i.e. two indicators intimately related to a housing shock, but also for the variation in financial assets. The variance decompositions for consumer durables confirms the heterogeneous pattern found in our graphical analysis. For the bottom 50% of households, the housing shock explains 44% (56%) of the common component of durable consumption using the small (large) dataset. Also for the poorest group, column four and seven show that the structural model explains between 20 and 25% of total variation in durable consumption. Although still considerable, the shock only explains about 20% (29%) of the common component and about 5% of total consumption variation for the richest percentile.

Our variance decomposition thus confirms that housing shocks are a vital component to model balance sheet dynamics in general and consumption dynamics of poorer households in particular. Table 1.2 also shows that common disturbances are overall able to capture a large fraction of fluctuations in balance sheet series.

Historical Decompositions

It is widely believed that the housing market has played a significant role in explaining the outbreak and strength of the Great Recession (Mian and Sufi, 2011, 2014). According to this narrative, negative house price shocks had a large effect on the economy via consumption spending. The aggregate collapse of consumption, so the argument, was mainly driven by poorer households, who cut spending disproportionally sharp as a result of declining housing wealth. In this section, we show historical decompositions to quantify how much the identified housing demand shock explains of the historically observed consumption patterns.

Series	Small Dataset VD	Small Dataset R^2	Small Dataset Total	Large Dataset VD	Large Dataset R^2	Large Dataset Total
Bottom 50%						
Total Net Wealth	0.19	0.29	0.06	0.24	0.30	0.07
Assets						
Real Estate Value	0.25	0.47	0.12	0.41	0.36	0.15
Consumer Durables	0.41	0.52	0.21	0.56	0.45	0.25
Financial	0.15	0.76	0.11	0.30	0.26	0.08
Liabilities						
Mortgages	0.31	0.50	0.15	0.53	0.43	0.22
Consumer Credit	0.23	0.30	0.07	0.39	0.22	0.08
50-90th Percentiles						
Total Net Wealth	0.23	0.52	0.12	0.42	0.33	0.14
Assets						
Real Estate Value	0.27	0.67	0.18	0.39	0.65	0.26
Consumer Durables	0.34	0.62	0.21	0.54	0.50	0.27
Financial	0.17	0.49	0.08	0.39	0.20	0.08
Liabilities						
Mortgages	0.29	0.64	0.19	0.45	0.65	0.29
Consumer Credit	0.22	0.33	0.07	0.39	0.265	0.10
90-99th Percentiles						
Total Net Wealth	0.14	0.91	0.13	0.41	0.29	0.12
Assets						
Real Estate Value	0.24	0.77	0.19	0.37	0.67	0.24
Consumer Durables	0.13	0.23	0.03	0.28	0.18	0.05
Financial	0.11	0.92	0.10	0.35	0.24	0.08
Liabilities						
Mortgages	0.38	0.49	0.19	0.59	0.40	0.24
Consumer Credit	0.27	0.26	0.07	0.46	0.21	0.10
Top 1%						
Total Net Wealth	0.15	0.92	0.14	0.35	0.36	0.12
Assets						
Real Estate Value	0.20	0.43	0.09	0.33	0.39	0.13
Consumer Durables	0.15	0.28	0.04	0.29	0.18	0.05
Financial	0.14	0.91	0.13	0.33	0.32	0.11
Liabilities						
Mortgages	0.37	0.59	0.22	0.59	0.55	0.33
Consumer Credit	0.13	0.20	0.03	0.22	0.13	0.03

Table 1.2. Variance Decompositions and R^2 for Balance Sheet Components. Notes: The second column reports the fraction of the common component of the series explained by the housing demand shock at a 32-quarter horizon. The third column shows the contribution of the common component to the overall variance of the indicators. The fourth column is simply the product of the previous two and represents the fraction of the total variability explained by the housing demand shock. The next three columns show the results for the large dataset.

Figure 1.5 attributes the historical variation in demeaned consumption of each wealth group to housing demand shocks. The shaded area marks the period of the Great Recession. Results from the large dataset are slightly more pronounced and relegated to the Appendix (see Figure 1.15). The bottom 50% of households cut durable consumption sharply during the crisis and stayed below the mean growth rate until 2017. The historical decomposition provides evidence that a large fraction of this contraction can be explained by housing demand shocks, both during and in the aftermath of the Great Recession. Although less pronounced, the cumulative effects of housing shocks also explain part of the cutback in durable consumption by the next 40%. In line with the variance decompositions, our structural model in general does not deliver much explanatory power for the evolution of consumption growth of the richest decile. During the Great Recession, however, the housing demand shock does explain about one percent of the contraction in consumption of the richest percentile. These results may seem contrasting to the findings of Mian and Sufi (2014), which suggest a muted reaction of the richest households' consumption when faced with negative housing wealth shocks. The apparent discrepancy is no contradiction, but just highlights the methodological difference: A housing demand shock is a housing wealth shock, but not exclusively. Consumption of each wealth group is arguably determined by distinct channels, and a housing demand shock impacts on all. A result that carries over from the partial to the general equilibrium setting, however, is that households belonging to the bottom 50% have strongly reduced consumption due to negative housing shocks. Importantly, these adverse effects have pushed growth below mean rates until only recently.

Long-term Elasticity of Consumption

The literature on wealth and collateral effects of housing frequently calculates the elasticity of consumption with respect to house price shocks. For example, Carroll et al. (2011) find a medium-term elasticity of 0.21 for total consumption and a one percent increase in house prices,[11] while Mian et al. (2013) report estimates between

[11] Since they calculate marginal propensities to consume (MPCs) in 2007 dollars, estimates have to be multiplied by the ratio of housing assets to personal consumption expenditures for 2007.

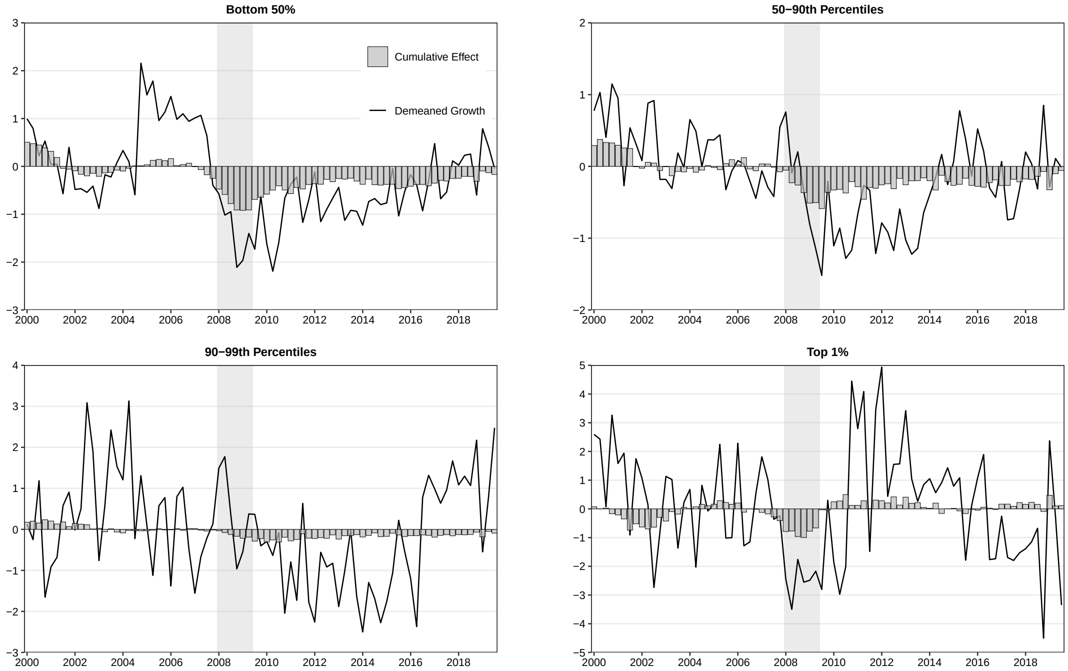


Figure 1.5: Historical contribution of housing demand shocks to durable consumption using the small dataset. Notes: Estimated on the small dataset, the historical decompositions show the median cumulative effect of housing demand shocks to demeaned growth of durable consumption for each wealth segment. The shaded area marks the period of the Great Recession according to the National Bureau's (NBER) business cycle dating committee.

0.11 and 0.29.[12] Based on their sensitivity instrument, Guren et al. (2021) arrive at somewhat lower estimates between 0.06 and 0.13. Impulse responses from the estimated DSGE model in Iacoviello and Neri (2010) translate into a first-year elasticity of total per capita consumption of around 0.07. Durable consumption estimates can be found in the online Appendix of Mian et al. (2013) and range from 0.2 to 0.43. Regarding heterogeneity across the wealth distribution, they provide evidence that elasticities of consumption are smaller for richer households. We compare these results to the long-term elasticities implied by our impulse responses for durable consumption and house prices. To obtain one aggregate price index, we take the

[12]Their estimates need to be scaled by the mean ratio of housing to total wealth in order to be comparable. The mean ratio in their data is between 0.25 and 0.33.

population-weighted average over the responses of the four (Census region) house price indices for each draw. Figure 1.6 shows the implied elasticities for all wealth groups based on the small dataset after 32 quarters. For each boxplot, the bold horizontal line indicates the posterior median. The first and third quartile is given by the upper and lower end of the box, respectively. The lower whisker shows the 16th percentile, while the upper whisker shows the 84th percentile of the posterior distribution. Figure 1.7 shows the results for the large dataset.

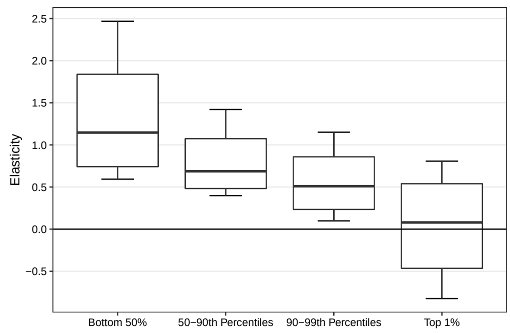


Figure 1.6: Small Dataset

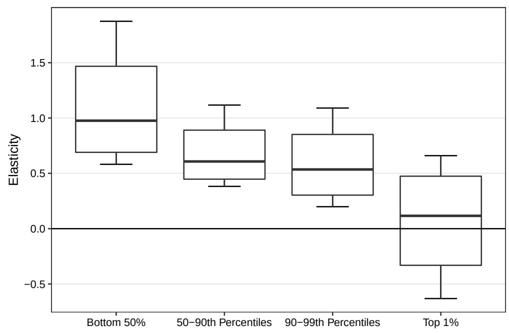


Figure 1.7: Large Dataset

Notes: The figure shows boxplots for the elasticity of consumption with respect to the average house price response based on the small and the large dataset. Boxplots are based on impulse responses after 32 quarters from the identified housing demand shock. For each boxplot, the bold horizontal line indicates the posterior median. The first and third quartile is given by the upper and lower end of the box, respectively. The lower whisker shows the 16% posterior percentile, while the upper whisker shows the 84% percentile.

The reported long-term median elasticities cover a wide range between 0.1 and 1.1. In line with the literature, we find estimates to be decreasing with household balance sheet strength. A one percent persistent increase of aggregate house prices is associated with a moderate 0.1% increase of durable consumption for the richest percentile. In contrast, consumption of the poorest 50% is estimated to match the rise in house prices one-to-one in the long-term. Median elasticities for households in the 50th to 90th wealth percentiles are between 0.6 and 0.7, which is close to the 0.5 estimates of the next 9%. The fact that median elasticities decline relatively slowly with wealth confirms the empirical results in Berger et al. (2017) and is at the center of recent work by Kaplan et al. (2014) on wealthy-hand-to-mouth households. Except for the richest percentile, the obtained elasticities are on the upper range when compared to the literature on wealth effects. Clearly, our median estimates need to be interpreted with caution, as posterior masses are wide for each boxplot. However, the location of the posterior distributions indicate the presence of heterogeneous elasticities across the wealth distribution.

1.5 Conclusion

In this article, we have reexamined the evidence on the propagation mechanism of housing demand shocks to aggregate economic activity and provide new evidence on the empirical link between aggregate outcomes and household balance sheet dynamics. Using a dynamic factor model and two different datasets, we identified shocks by imposing theoretically motivated sign restrictions on a few selected economic indicators and a large number of housing series.

The results show that a positive housing demand shock generates a substantial and persistent economic boom. In line with the outlined transmission channels, we find a large increase in residential investment, aggregate mortgage loans and aggregate consumption. Real economic indicators such as GDP, the industry capacity utilization rate, and the industrial production index increase despite an immediate rise of interest rates. Variance decompositions reveal that housing demand shocks explain a large fraction of business cycle fluctuations in general and house price variation in particular. The common factors are shown to adequately capture the movements in observable variables.

The second part of our analysis provides evidence that housing demand shocks are key to understand household balance sheet dynamics. Based on detailed time series for four different wealth percentile groups, our results support the finding that heterogeneity across the wealth distribution matters. While this is true especially for durable consumption, we also find important cross-sectional similarities: All wealth segments profit from a positive housing demand shock in terms of net wealth. Also, each wealth group extracts a significant amount of equity from housing, although the response is most pronounced for the poorest 50% of households. Given this groups' substantial holdings of consumer credit, we might expect that the new liquidity is used to pay down the more expensive consumer credit balances. In line with the literature, we do not find such an effect. Consumer credit increases for all groups but the richest percentile, which holds a negligible amount of consumer credit and is likely not to be drained from the associated costs. Importantly, the response of durable consumption is a decreasing function of net wealth. Variance decompositions attribute a substantial fraction of durable consumption variation to the housing demand shock, especially for the bottom 50% of households. The structural model is in general well suited to explain household balance sheet variation. A historical decomposition of durable consumption confirms the narrative that housing shocks have caused a significant part of the fall in durable consumption for the poor during the Great Recession. However, while the cumulative effects in general are negligible for the richest percentile of households, we find evidence that the shock has negatively affected their consumption in that period. Finally, we translate impulse responses into elasticities of consumption with respect to house prices. Although uncertainty regarding these estimates is high, results suggest a larger spending sensitivity for weaker balance sheet households. Altogether, our findings indicate that household balance sheets play a significant role in determining aggregate outcomes. All results are robust to two different datasets and model specifications.

1.A Appendix

1.A.1 Algorithm

Conditional on the factors, we assume the parameters of the state equation to follow Jeffreys prior. In the observation equation the covariance matrix R is assumed to be diagonal with the loadings and R following a normal-inverse gamma prior. So given the $((T-s) \times m(s+1))$ matrix F and Y containing the factors and the data, each row of Λ^* and the associated diagonal entry of R is drawn equation-by-equation. If we denote the ith row of the loadings matrix Λ^* by Λ_i^*, y_i the $((T-s) \times 1)$ vector consisting of the ith observation variable, and σ_i^2 the ith diagonal entry in R, then the priors are specified as $\pi(\sigma_i^2) = IG(v_0, s_0)$ and $\pi(\Lambda_i^{*\prime}|\sigma_i^2) = N(\Lambda_{i,0}^{*\prime}, \sigma_i^2 \Sigma_{\lambda,0}^{-1})$. In this case the posterior is characterized as follows

$$\begin{aligned}
&\Lambda_i^{*\prime}|\sigma_i^2, y_i, F \sim N(\Lambda_{i,T}^{*\prime}, \sigma_i^2(F'F + \Sigma_{\lambda,0})^{-1}) \\
&\text{where } \Lambda_{i,T}^{*\prime} = (F'F + \Sigma_{\lambda,0})^{-1}(F'F\hat{\Lambda}_i^{*\prime} + \Sigma_{\lambda,0}\Lambda_{i,0}^{*\prime}) \\
&\text{and } \hat{\Lambda}_i^{*\prime} = (F'F)^{-1}F'y_i \\
&\sigma_i^2|y_i, F \sim IG(v_T, s_T) \\
&\text{where } v_T = \frac{T-s+v_0}{2} \\
&\text{and } s_T = \frac{(y_i - F\Lambda_{i,T}^{*\prime})'(y_i - F\Lambda_{i,T}^{*\prime}) + (\Lambda_{i,T}^{*\prime} - \Lambda_{i,0}^{*\prime})'\Lambda_{i,0}^{*\prime}(\Lambda_{i,T}^{*\prime} - \Lambda_{i,0}^{*\prime}) + s_0}{2}
\end{aligned} \tag{1.3}$$

Following Bai and Wang (2015) the upper left $(m \times m)$ block of Λ^* is restricted to be an identity matrix. The state equation can be written in matrix form as

$$\xi = X\Phi^{*\prime} + U \tag{1.4}$$

where ξ is a $((T-h) \times m)$ matrix containing the factors, X is a $((T-h) \times mh)$ matrix containing the lagged factors, and Φ^* is equal to the upper m rows of Φ. Defining $\phi = vec(\Phi^{*\prime})$ and employing a diffuse prior $\pi(Q, \phi) \propto \pi(Q) \propto |Q|^{-(m+1)/2}$ gives us a posterior in the form of

$$
\begin{aligned}
&\phi|Q,\xi,X \sim N(\hat{\phi}, Q \otimes (X'X)^{-1}) \\
&\text{where } \hat{\phi} = vec((X'X)^{-1}X'\xi) \\
&Q|\xi,X \sim IW(\hat{U}'\hat{U}, T - mh) \\
&\text{where } \hat{U} = \xi - X(X'X)^{-1}X'\xi
\end{aligned}
\tag{1.5}
$$

The priors for the observation equation are set to be $v_0 = m + 3$, $s_0 = 0.1$, $\lambda_0^{*\prime} = 0_{m(s+1)\times 1}$, while $\Sigma_{\lambda,0} = I_{m(s+1)}$. The factors are provided by the Kalman filter and subsequent backward sampling. The Gibbs sampler iterates between drawing the factors and drawing the parameters, such that the joint posterior is numerically approximated.

1.A.2 Impulse Responses

Benchmark Model

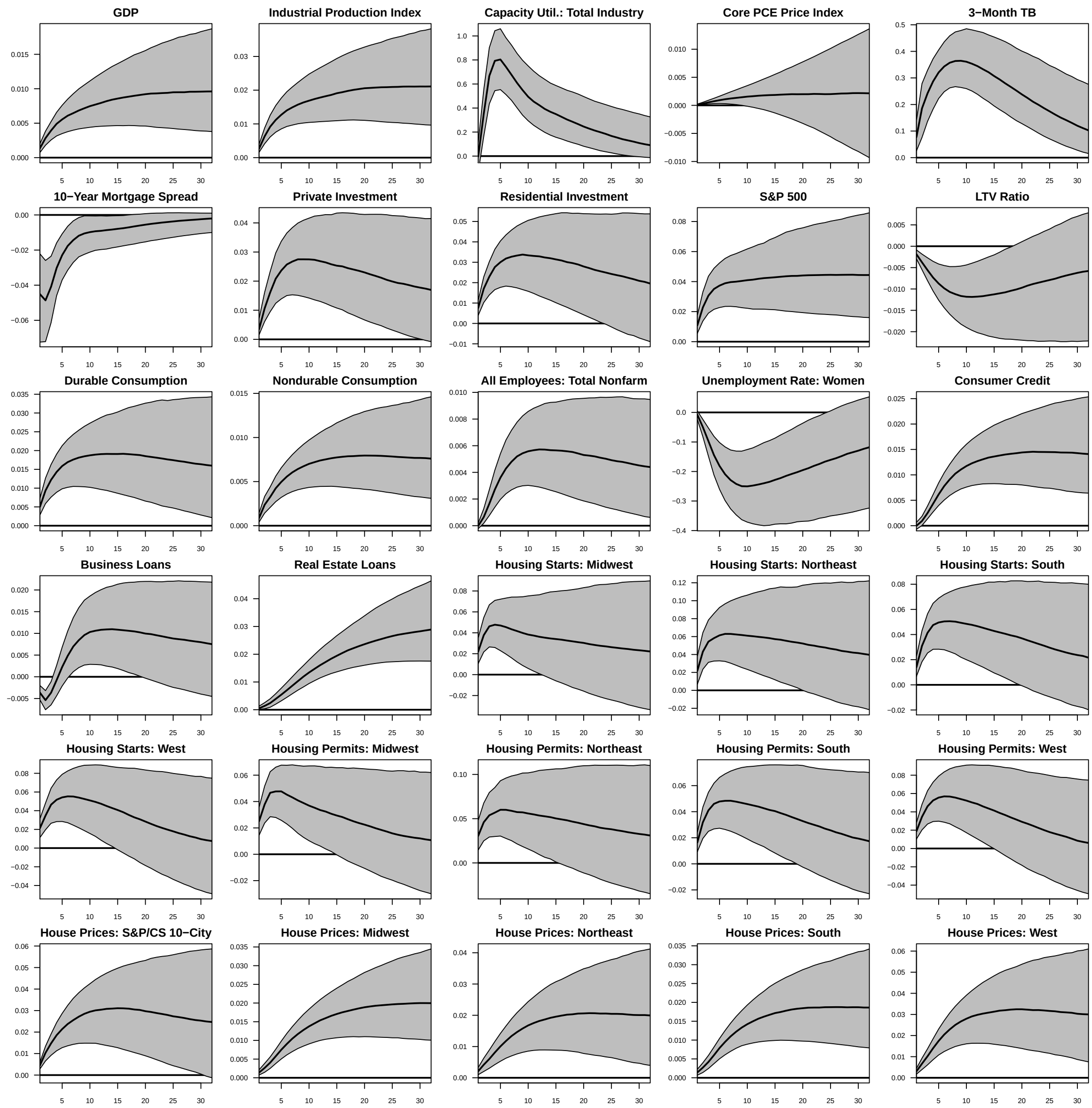


Figure 1.8: Dynamic responses of variables to housing demand shock using the large dataset. Notes: The figure plots the impulse responses of the level of key variables to the identified housing demand shock. The gray areas indicate the 68% posterior probability regions. The straight black line indicates the posterior median at each horizon.

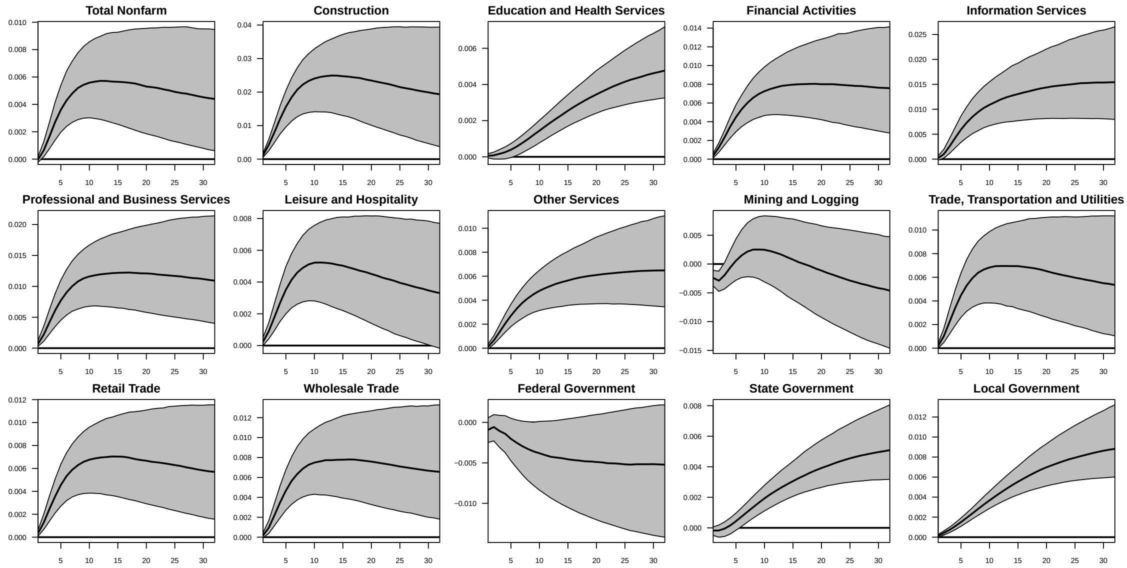


Figure 1.9: Dynamic responses of sectoral employment variables to housing demand shock using the large dataset. Notes: The figure plots the impulse responses of the level of sectoral employment variables to the identified housing demand shock. The gray areas indicate the 68% posterior probability regions. The straight black line indicates the posterior median at each horizon.

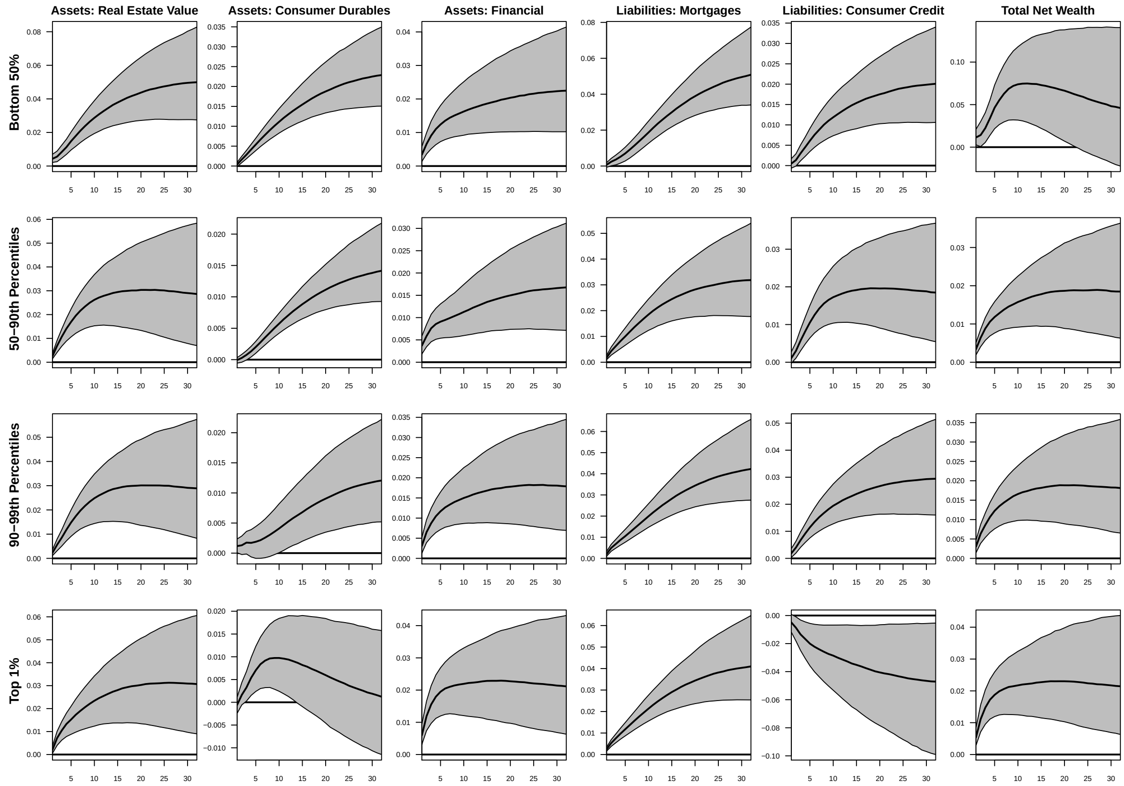


Figure 1.10: Dynamic responses of household balance sheet variables to housing demand shock using the large dataset. Notes: The figure plots the impulse responses of the level of households' balance sheet components to the identified housing demand shock. The gray areas indicate the 68% posterior probability regions. The straight black line indicates the posterior median at each horizon.

Robustness of Results across Model Specifications

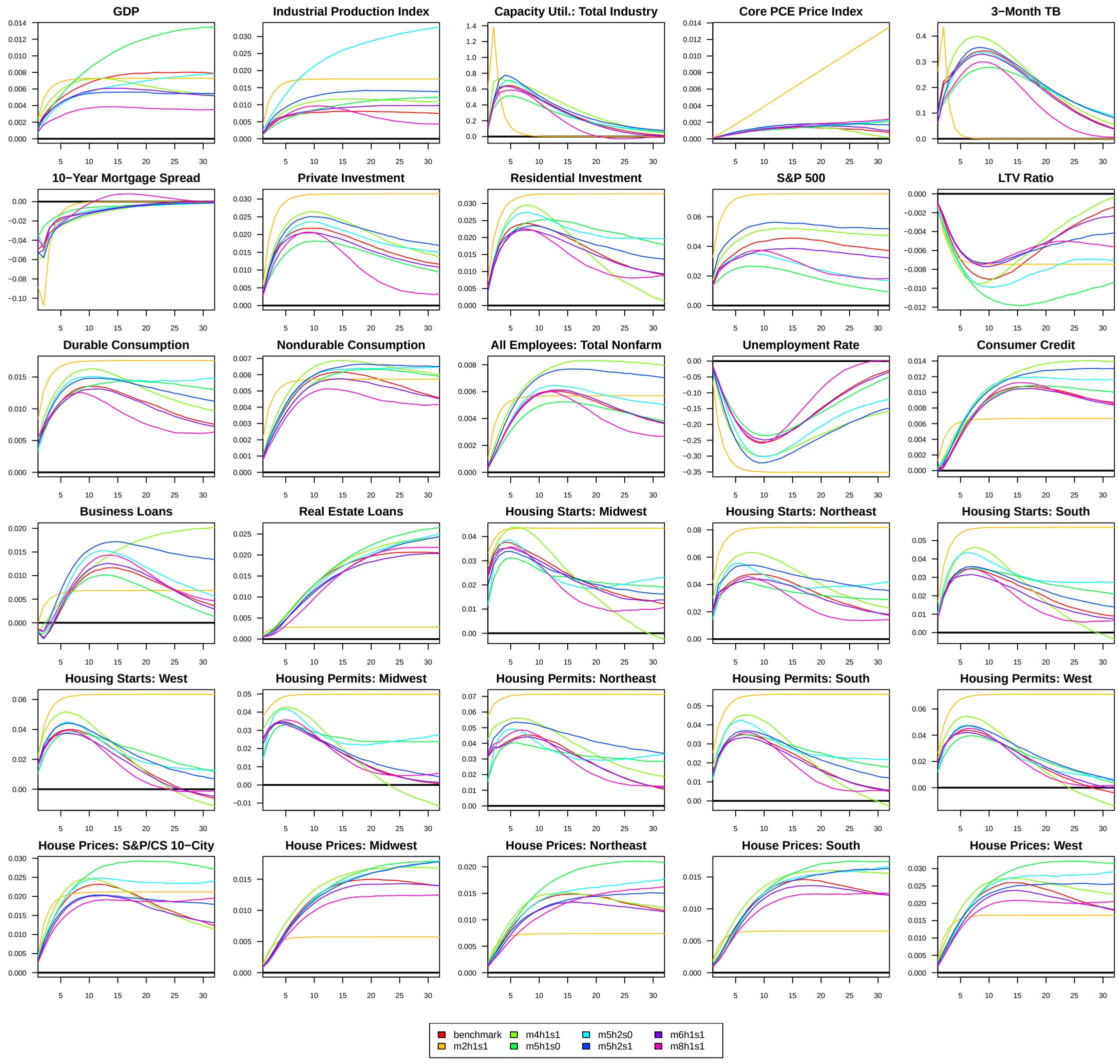


Figure 1.11: Dynamic median responses of variables to housing demand shock using the small dataset and various model specifications. Notes: The figure plots the median impulse responses of the level of key variables to the identified housing demand shock for a wide range of model specifications.

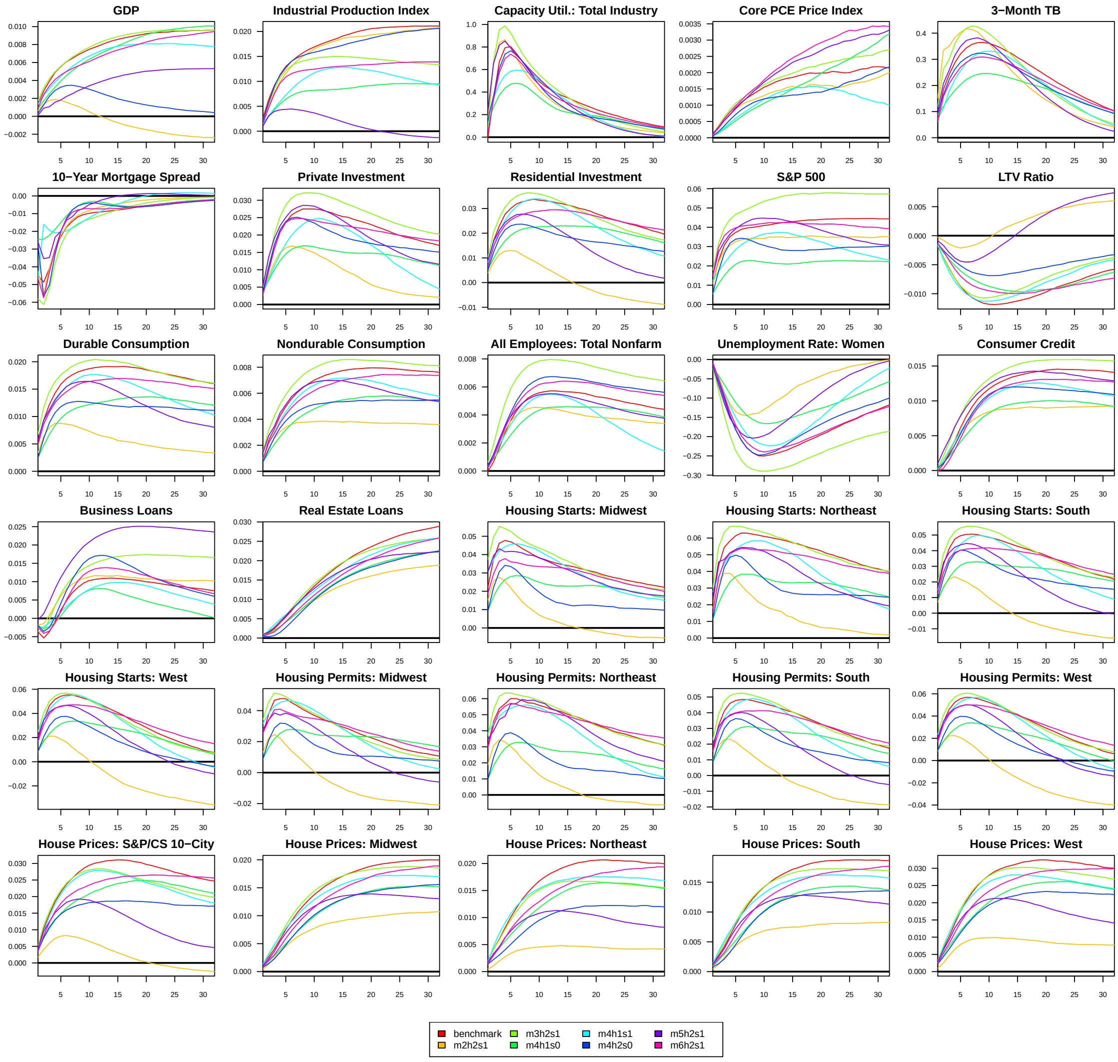


Figure 1.12: Dynamic median responses of variables to housing demand shock using the large dataset and various model specifications. Notes: The figure plots the median impulse responses of the level of key variables to the identified housing demand shock for a wide range of model specifications.

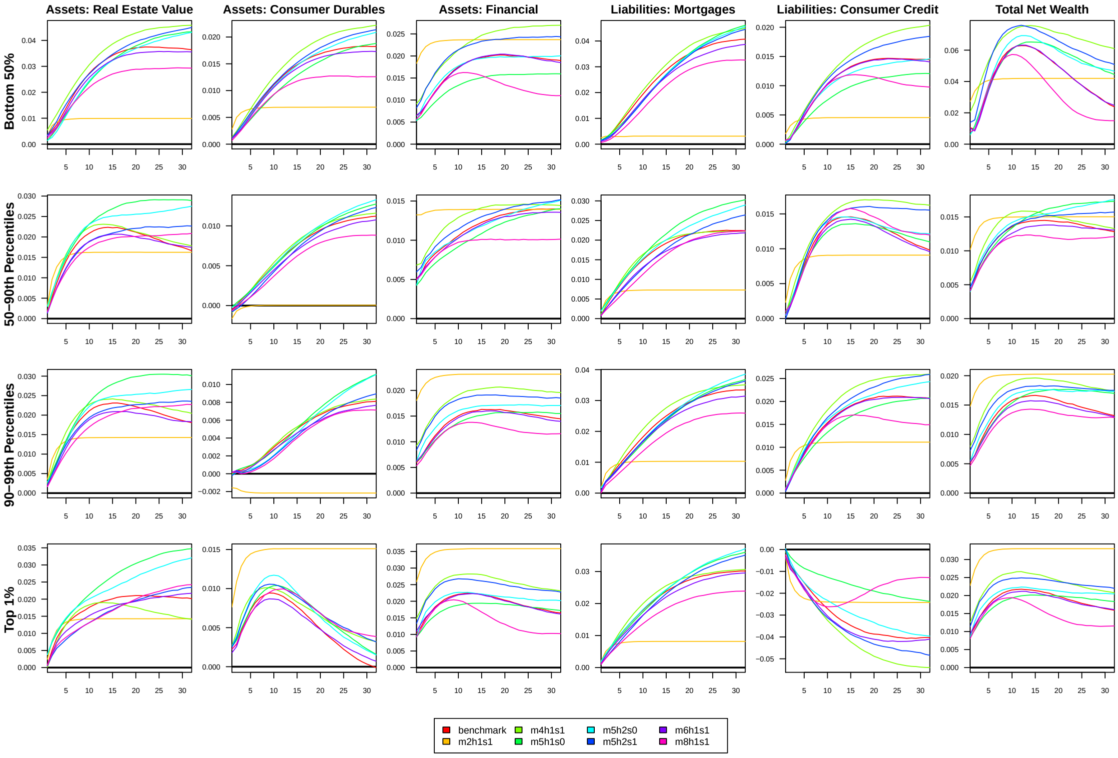


Figure 1.13: Dynamic median responses of household balance sheet variables to housing demand shock using the small dataset and various model specifications. Notes: The figure plots the median impulse responses of the level of households' balance sheet components to the identified housing demand shock for a wide range of model specifications.

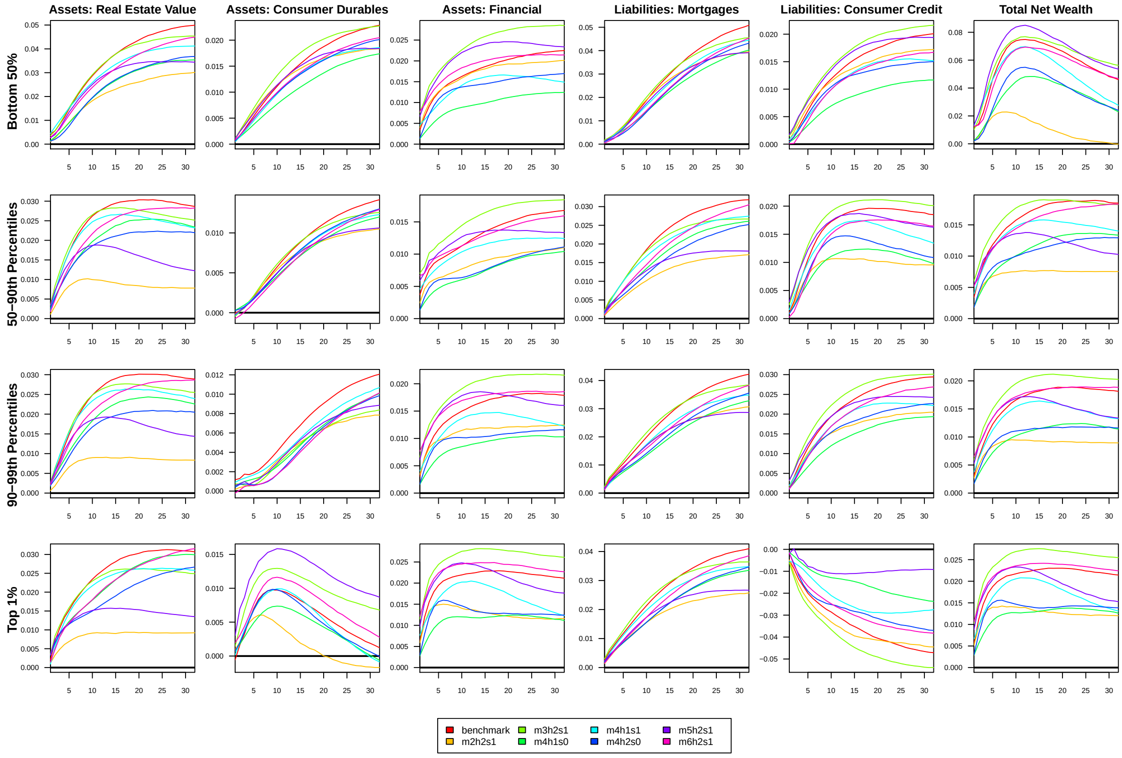


Figure 1.14: Dynamic median responses of household balance sheet variables to housing demand shock using the large dataset and various model specifications. Notes: The figure plots the median impulse responses of the level of households' balance sheet components to the identified housing demand shock for a wide range of model specifications.

1.A.3 Historical Decompositions

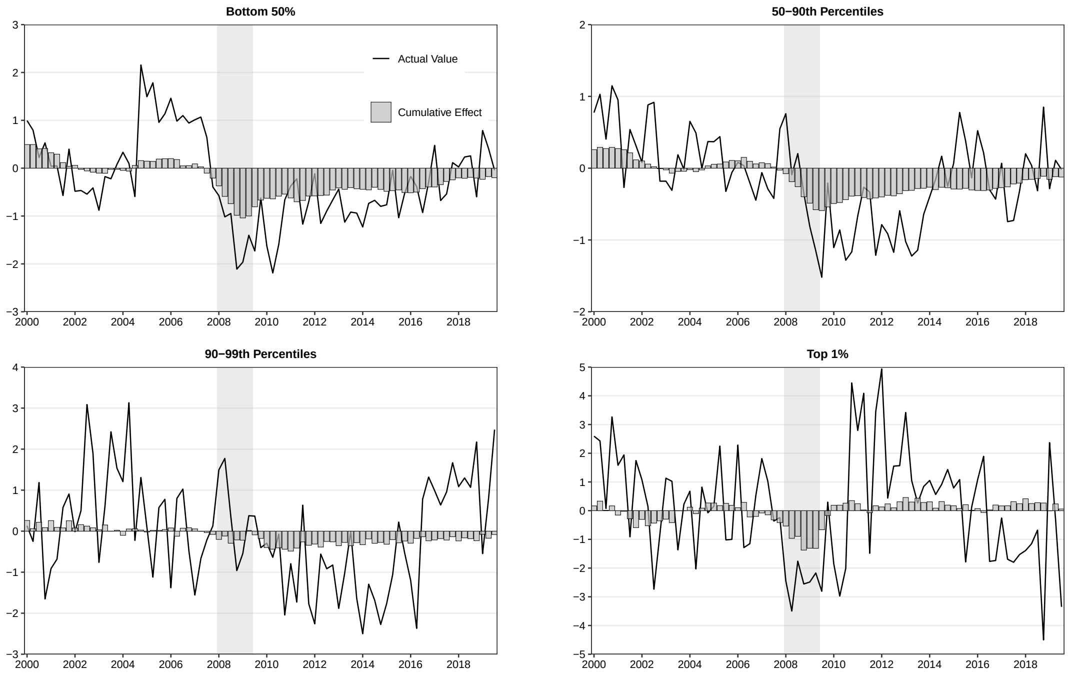


Figure 1.15: Historical contribution of housing demand shocks to durable consumption using the large dataset. Notes: Estimated on the large dataset, the historical decompositions show the median cumulative effect of housing demand shocks to demeaned growth of durable consumption for each wealth segment. The shaded area marks the period of the Great Recession according to the National Bureau's (NBER) business cycle dating committee.

1.A.4 Data

The first column gives the variable names from the FRED-QD database (see McCracken and Ng, 2020), except for the series ending with *. For these cases see the description in the second column. The third column gives the code of how each series y_i is transformed to stationarity. Here, (1) means no transformation, (2) is Δy_{it}, (5) is $\Delta log(y_{it})$, and (6) is $\Delta^2 log(y_{it})$. The fourth column indicates whether the series is part of the large dataset (0), small and large dataset (1), or only the small dataset (2).

Data Group 1: NIPA

Variable	Description	Transformation	Dataset
GDPC1	Real GDP	5	1
PCECC96	Real Consumption	5	1
PCDGx	Real Durable Consumption	5	1
PCESVx	Real Service Consumption	5	1
PCNDx	Real Nondurable Consumption	5	1
GPDIC1	Real Gross Private Domestic Investment	5	1
Y033RC1Q027SBEAx	Real Gross Private Domestic Fixed Investment: Equipment	5	1
PNFIx	Real Private Fixed Investment: Nonresidential	5	1
PRFIx	Real Private Fixed Investment: Residential	5	1
A014RE1Q156NBEA	Shares of GDP: Gross Private Domestic Investment: Private Inventories (Percent)	1	1
A823RL1Q225SBEA	Real Federal Government Consumption and Gross Investment (Percent Change)	1	1
FGRECPTx	Real Federal Government Current Receipts	5	0
SLCEx	Real State and Local Government Consumption	5	0
EXPGSC1	Real Exports	5	0
IMPGSC1	Real Imports	5	0

Data Group 2: Industrial Production

Variable	Description	Transformation	Dataset
INDPRO	Industrial Production Index	5	1
IPDMAT	Industrial Production Index: Durable Materials	5	0
IPNMAT	Industrial Production Index: Nondurable Materials	5	0
IPDCONGD	Industrial Production Index: Durable Consumer Goods	5	0
IPB51110SQ	Industrial Production Index: Durable Automotive Products	5	0
IPNCONGD	Nondurable Consumer Goods	5	0
IPBUSEQ	Business Equipment	5	0
IPB51220SQ	Consumer Energy Products	5	0
TCU	Capacity Utilization: Total Industry (Percent)	1	1
CUMFNS	Capacity Utilization: Manufacturing (Percent)	1	1

Data Group 3: Employment and Unemployment

Variable	Description	Transformation	Dataset
PAYEMS	All Employees: Total Nonfarm	5	1
USPRIV	All Employees: Total Private Industries	5	2
MANEMP	All Employees: Manufacturing	5	2
DMANEMP	All Employees: Durable Goods	5	0
USCONS	All Employees: Construction	5	0
USEHS	All Employees: Education and Health Services	5	0
USFIRE	All Employees: Financial Activities	5	0
USINFO	All Employees: Information Services	5	0
USPBS	All Employees: Professional and Business Services	5	0
USLAH	All Employees: Leisure and Hospitality	5	0
USSERV	All Employees: Other Services	5	0
USMINE	All Employees: Mining and Logging	5	0
USTPU	All Employees: Trade, Transportation and Utilities	5	0
USTRADE	All Employees: Retail Trade	5	0
USWTRADE	All Employees: Wholesale Trade	5	0
CES9091000001	All Employees: Federal Government	5	0
CES9092000001	All Employees: State Government	5	0
CES9093000001	All Employees: Local Government	5	0
UNRATE	Civilian Unemployment Rate (Percent)	2	2
LNS14000012	Unemployment Rate - 16 to 19 years (Percent)	2	0
LNS14000025	Unemployment Rate Men - 20 years and Over (Percent)	2	0
LNS14000026	Unemployment Rate Women - 20 years and Over (Percent)	2	0
UEMPLT5	Number of Civilians Unemployed - Less than 5 Weeks	5	0
UEMP5TO14	Number of Civilians Unemployed - 5 to 14 Weeks	5	0
UEMP15T26	Number of Civilians Unemployed - 15 to 26 Weeks	5	0
UEMP27OV	Number of Civilians Unemployed - More than 27 Weeks	5	0
LNS13023621	Unemployment Level - Job Losers	5	0
LNS13023557	Unemployment Level - Reentrants to Labor Force	5	0
LNS13023705	Unemployment Level - Job Leavers	5	0
LNS13023569	Unemployment Level - New Entrants	5	0
LNS12032194	Employment Level - Part-Time for Economic Reasons	5	0
AWHMAN	Average Weekly Hours of Production and Nunsupervisory Employees: Manufacturing	1	0
AWHNONAG	Average Weekly Hours of Production and Nunsupervisory Employees: Total Private	2	0
AWOTMAN	Average Weekly Overtime Hours of Production and Nunsupervisory Employees: Manufacturing	2	0

Data Group 4: Housing

Variable	Description	Transformation	Dataset
HOUST	Total Housing Starts: New Privately Owned Housing Units	5	1
HOUST5F	Housing Starts: New Privately Owned 5 or More Unit Structures	5	1
PERMIT	New Private Housing Units Authorized by Building Permits	5	1
HOUSTMW	Housing Starts in Midwest Census Region	5	1
HOUSTNE	Housing Starts in Northeast Census Region	5	1
HOUSTS	Housing Starts in South Census Region	5	1
HOUSTW	Housing Starts in West Census Region	5	1
midwest_hp*	FRED: Average of All-Transactions House Price Indices for the East North Central and West North Central Census Division	5	1
northeast_hp*	FRED: Average of All-Transactions House Price Indices for the New England and Middle Atlantic Census Division	5	1
south_hp*	FRED: Average of All-Transactions House Price Indices for the South Atlantic, East South Central and West South Central Census Division	5	1
west_hp*	FRED: Average of All-Transactions House Price Indices for the Mountain and Pacific Census Division	5	1
SPCS10RSA	S&P/Case-Shiller 10-City Composite Home Price Index	5	1
PERMITNE	New Private Housing Units Authorized by Building Permits in Northeast Census Region	5	1
PERMITMW	New Private Housing Units Authorized by Building Permits in Midwest Census Region	5	1
PERMITS	New Private Housing Units Authorized by Building Permits in South Census Region	5	1
PERMITW	New Private Housing Units Authorized by Building Permits in West Census Region	5	1

Data Group 5: Inventories, Orders, and Sales

Variable	Description	Transformation	Dataset
RSAFSx	Real Retail and Food Services Sales	5	0
AMDMNOx	Real Manufacturers' New Orders: Durable Goods	5	0
AMDMUOx	Real Value of Manufacturers' Unfilled Orders: Durable Goods Industries	5	0
ANDENOx	Real Value of Manufacturers' New Orders for Capital Goods: Nondefense Capital Goods Industries	5	0
INVCQRMTSPL	Real Manufacturing and Trade Inventories	5	0

Data Group 6: Prices

Variable	Description	Transformation	Dataset
PCECTPI	Personal Consumption Expenditures Price Index	6	1
PCEPILFE	Personal Consumption Expenditures Price Index excl. Food and Energy	6	1
GDPCTPI	Gross Domestic Product Price Index	6	2
GPDICTPI	Gross Private Domestic Investment Price Index	6	0
IPDBS	Business Sector: Implicit Price Deflator	6	0
DMOTRG3Q086SBEA	Personal Consumption Expenditures Price Index: Durable Goods: Motor Vehicles and Parts	6	0
DFDHRG3Q086SBEA	Personal Consumption Expenditures Price Index: Durable Goods: Furnishings and Durable Household Equipment	6	0
DREQRG3Q086SBEA	Personal Consumption Expenditures Price Index: Durable Goods: Recreational Goods and Vehicles	6	0
DODGRG3Q086SBEA	Personal Consumption Expenditures Price Index: Durable Goods: Other Durable Goods	6	0
DFXARG3Q086SBEA	Personal Consumption Expenditures Price Index: Nondurable Goods: Food and Beverages	6	0
DCLORG3Q086SBEA	Personal Consumption Expenditures Price Index: Nondurable Goods: Clothing and Footwear	6	0
DGOERG3Q086SBEA	Personal Consumption Expenditures Price Index: Nondurable Goods: Gasoline and Other Energy Goods	6	0
DONGRG3Q086SBEA	Personal Consumption Expenditures Price Index: Nondurable Goods: Other Nondurable Goods	6	0
DHUTRG3Q086SBEA	Personal Consumption Expenditures Price Index: Services: Housing and Utilities	6	0
DHLCRG3Q086SBEA	Personal Consumption Expenditures Price Index: Services: Health Care	6	0
DTRSRG3Q086SBEA	Personal Consumption Expenditures Price Index: Services: Transportation	6	0
DRCARG3Q086SBEA	Personal Consumption Expenditures Price Index: Services: Recreation Services	6	0
DFSARG3Q086SBEA	Personal Consumption Expenditures Price Index: Services: Food and Accommodations	6	0
DIFSRG3Q086SBEA	Personal Consumption Expenditures Price Index: Services: Finance and Insurance	6	0
DOTSRG3Q086SBEA	Personal Consumption Expenditures Price Index: Services: Other Services	6	0
WPSFD49502	Producer Price Index by Commodity for Finished Goods	6	0
WPSFD4111	Producer Price Index by Commodity for Finished Consumer Goods	6	0
PPIIDC	Producer Price Index by Commodity Industrial Commodities	6	0
WPSID61	Producer Price Index by Commodity Intermediate Materials: Supplies and Components	6	0
WPU0531	Producer Price Index by Commodity for Fuels and Related Products and Power: Natural Gas	5	0
WPU0561	Producer Price Index by Commodity for Fuels and Related Products and Power: Crude Petroleum	5	0

Data Group 7: Earnings and Productivity

Variable	Description	Transformation	Dataset
COMPRMS	Manufacturing Sector: Real Compensation Per Hour	5	0
COMPRNFB	Nonfarm Business Sector: Real Compensation Per Hour	5	0
RCPHBS	Business Sector: Real Compensation Per Hour	5	0
OPHMFG	Manufacturing Sector: Real Output Per Hour of All Persons	5	0
OPHNFB	Nonfarm Business Sector: Real Output Per Hour of All Persons	5	0
ULCMFG	Manufacturing Sector: Unit Labor Cost	5	0
ULCNFB	Nonfarm Business Sector: Unit Labor Cost	5	0
UNLPNBS	Nonfarm Business Sector: Unit Nonlabor Payments	5	0

Data Group 8: Interest Rates

Variable	Description	Transformation	Dataset
FEDFUNDS	Effective Federal Funds Rate	1	1
TB3MS	3-Month Treasury Bill: Secondary Market Rate	1	1
MORTGAGE30US	30-Year Conventional Mortgage Rate	1	1
BAA10YM	Moody‚Äôs Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity	1	0
MORTG10YRx	30-Year Conventional Mortgage Rate Relative to 10-Year Treasury Constant Maturity	1	1
TB6M3Mx	6-Month Treasury Bill Minus 3-Month Treasury Bill	1	0
GS1TB3Mx	1-Year Treasury Constant Maturity Minus 3-Month Treasury Bill	1	0
GS10TB3Mx	10-Year Treasury Constant Maturity Minus 3-Month Treasury Bill	1	0
CPF3MTB3Mx	3-Month Commercial Paper Minus 3-Month Treasury Bill	1	0
ShadowRate*	Quantifies the stance of monetary policy in zero lower bound environment and characterizes the term structure of interest rates (see Wu and Xia, 2016)	1	1

Data Group 9: Credit

Variable	Description	Transformation	Dataset
BUSLOANSx	Real Commercial and Industrial Loans, All Commercial Banks	5	1
NONREVSLx	Total Outstanding Real Nonrevolving Credit Owned and Securitized	5	1
REALLNx	Real Real Estate Loans, All Commercial Banks	5	1
REVOLSLx	Total Real Revolving Credit Owned and Securitized, Outstanding	5	1
TOTALSLx	Total Consumer Credit Outstanding	5	1
DRIWCIL	FRB Senior Loans Officer Opinions. Net Percentage of Domestic Respondents Reporting Increased Willingness to Make Consumer Installment Loans	1	0
LTV*	Loan-to-Value Ratio. FRED Variables: Total Credit to Private Non-Financial Sector, Adjusted for Breaks; and Households and Nonprofit Organizations: Real Estate at Market Value	2	1

Data Group 10: Exchange Rates

Variable	Description	Transformation	Dataset
TWEXAFEGSMTHx	Trade Weighted U.S. Dollar Index: Major Currencies	5	0
EXSZUSx	Switzerland / U.S. Foreign Exchange Rate	5	0
EXJPUSx	Japan / U.S. Foreign Exchange Rate	5	0
EXUSUKx	U.S. / U.K. Foreign Exchange Rate	5	0
EXCAUSx	Canada / U.S. Foreign Exchange Rate	5	0

Data Group 11: Other

Variable	Description	Transformation	Dataset
UMCSENTx	University of Michigan: Consumer Sentiment Index	1	0
USEPUINDXM	Economic Policy Uncertainty Index for United States	2	0

Data Group 12: Stock Markets

Variable	Description	Transformation	Dataset
S&P 500	S&P,Äôs Common Stock Price Index: Composite	5	1
VXOCLSx	CBOE S&P 100 Volatility Index: VXO	1	0

Data Group 13: Household Balance Sheet by Wealth

Variable	Description	Transformation	Dataset
real_estate_bottom50*	FRED: Real Estate Held by the Bottom 50%	5	1
mortgages_bottom50*	FRED: Mortgages Held by the Bottom 50%	5	1
consumer_durables_bottom50*	FRED: Consumer Durables Held by the Bottom 50%	5	1
consumer_credit_bottom50*	FRED: Consumer Credit Held by the Bottom 50%	5	1
financial_bottom50*	FRED: Financial Assets Held by the Bottom 50%	5	1
total_net_bottom50*	FRED: Total Net Worth Held by the Bottom 50%	5	1
real_estate_50to90th*	FRED: Real Estate Held by the 50th to 90th Wealth Percentiles	5	1
mortgages_50to90th*	FRED: Mortgages Held by the 50th to 90th Wealth Percentiles	5	1
consumer_durables_50to90th*	FRED: Consumer Durables Held by the 50th to 90th Wealth Percentiles	5	1
consumer_credit_50to90th*	FRED: Consumer Credit Held by the 50th to 90th Wealth Percentiles	5	1
financial_50to90th*	FRED: Financial Assets Held by the 50th to 90th Wealth Percentiles	5	1
total_net_50to90th*	FRED: Total Net Worth Held by the 50th to 90th Wealth Percentiles	5	1
real_estate_90to99th*	FRED: Real Estate Held by the 90th to 99th Wealth Percentiles	5	1
mortgages_90to99th*	FRED: Mortgages Held by the 90th to 99th Wealth Percentiles	5	1
consumer_durables_90to99th*	FRED: Consumer Durables Held by the 90th to 99th Wealth Percentiles	5	1
consumer_credit_90to99th*	FRED: Consumer Credit Held by the 90th to 99th Wealth Percentiles	5	1
financial_90to99th*	FRED: Financial Assets Held by the 90th to 99th Wealth Percentiles	5	1
total_net_90to99th*	FRED: Total Net Worth Held by the 90th to 99th Wealth Percentiles	5	1
real_estate_top1*	FRED: Real Estate Held by the Top 1%	5	1
mortgages_top1*	FRED: Mortgages Held by the Top 1%	5	1
consumer_durables_top1*	FRED: Consumer Durables Held by the Top 1%	5	1
consumer_credit_top1*	FRED: Consumer Credit Held by the Top 1%	5	1
financial_top1*	FRED: Financial Assets Held by the Top 1%	5	1
total_net_top1*	FRED: Total Net Worth Held by the Top 1%	5	1

Notes: To ensure comparability between the household balance sheet series we eliminate population differences by dividing the series of each group by the respective number of population percentiles. All household balance sheet series are deflated using the core PCE deflator.

Chapter 2

The Reconciled Output Gap: A State-Space Framework to Model Revisions[1]

2.1 Introduction

Fiscal policymakers rely on precise estimates of the output gap – the difference between actual and potential output – to determine the cyclical position of the economy and to detect structural imbalances. The output gap is also an important determinant in central banks' reaction functions and is generally assumed to be a predictor of inflationary or disinflationary pressure (see, for instance, Gerlach and Smets, 1999; Sturm and De Haan, 2011; Coibion and Gorodnichenko, 2015). Recent years have seen the introduction of structural budget balances as explicit policy targets in several countries.[2] The output gap is also used in financial regulation frameworks to estimate countercyclical capital buffers (Drehmann et al., 2010). A

[1]This chapter is joint work Florian Eckert and Nina Mühlebach

[2]The European Fiscal Stability Treaty of 2012, for instance, obligates member states to pursue a fiscal policy such that structural deficits do not exceed a country-specific medium-term budgetary objective. The Swiss federal government has been using output gap estimates to determine an expenditure ceiling since 2004, with the aim of achieving a structurally balanced budget.

timely and reliable decomposition of output into a trend and a cyclical component is therefore crucial.

Despite its importance, estimating the output gap remains a challenge. Unobservable by nature, the results often depend on the underlying model assumptions and are therefore, to a certain extent, at the researcher's discretion. The credibility of rule-based fiscal planning can be compromised if policy decisions depend on output gap estimates that are susceptible to manipulation. Furthermore, many existing methods produce estimates that are subject to significant fluctuations, especially at the end of the sample (Orphanides and van Norden, 2002; Orphanides, 2003a). Revisions in the output gap can be caused by revisions in the underlying data, parameter instabilities, or end-point instabilities. A method suffers from end-point instabilities if it is sensitive to additional information at the current edge, i.e., the most recent gap estimates are substantially revised when a new observation is added to the time series (Barbarino et al., 2020). Erratic output gaps can pose a problem for policymakers that rely on timely and accurate estimates as input for their models (see, for instance, Kuttner, 1992; Smets, 2002; Orphanides and van Norden, 2005). It is further complicated by the fact that many estimates of potential output growth react to transitory shocks (Coibion et al., 2018). As a result, institutions that are concerned with cyclically adjusted budget balances and long-run fiscal trends may base their decisions on an unstable foundation.

We propose a state-space model to combine output gap estimates from multiple sources and vintages. Following the reconciliation framework of Jacobs et al. (2022), we use multiple data releases to distinguish between noise and news measurement errors. Therefore, this article contributes to the literature on measurement reconciliation and potential output estimation in various ways. First, we extend the error decomposition framework in Jacobs et al. (2022) from the reconciliation of expenditure and income side estimates of GDP to the case where multiple output gap estimates are reconciled. Second, we treat the output of various models as noisy estimates of the unobserved cyclical position, which are subject to revisions as new information becomes available. We provide a 'true' output gap in the sense that it is the best linear unbiased estimate. Our method is also able to deal with ragged edges and limited historical availability of data. Third, we provide a comprehensive real-time evaluation of our proposed reconciliation model, competing combination methods, and a large number of well-established models, reexamining Orphanides and van Norden (2002). We show that our measure is resilient to revisions in real-time in general but also in times of economic crisis. Furthermore, we evaluate the underlying output gap methods regarding their information content for the reconciled output gap and decompose their revisions into news and noise errors.

The uncertainty surrounding data, models, or parameters is often addressed by combining information from multiple sources. There is, for instance, substantial evidence for the benefits of model combination in forecasting (see Stock and Watson, 2004; Timmermann, 2006; Wickramasuriya et al., 2019, and references therein). Aruoba et al. (2016) and Jacobs et al. (2022) combine information from expenditure and income-side estimates of US GDP growth to extract a robust measure of output. However, research on the benefits of combining output gap estimates has remained limited so far and is complicated by several factors. Garratt et al. (2014) construct ensemble nowcasts for inflation using a linear opinion pool of several output gap estimation methods. The weight of each method is derived from its predictive accuracy for inflation. In a similar approach, Morley and Piger (2012) combine several cycles using posterior model probabilities as weights. Pichette et al. (2019) provide evidence that the mean and median of several output gap measures improve forecasts of core inflation in Canada. Similarly, Guérin et al. (2015) show that using the average of several output gap estimates improves the reliability of potential output estimates in real time and softens the impact of data revisions. Grant and Chan (2017) reconcile measures by embedding the Hodrick-Prescott (HP) filter into an unobserved components model, showing that it fits the data better than a regular HP filter. While these approaches deal with the issue of measurement errors in the underlying models, they do not consider how much these models are subject to revisions. Hence, these combination approaches do not necessarily improve the reliability of estimates at the end of the sample.

The remainder of the paper is structured as follows. Section 2.2 presents the reconciliation framework and discusses the assumptions necessary to identify noise and news measurement errors. Section 2.3 presents the data and the underlying models. Section 2.4 discusses our reconciled output gap estimate using the full sample. Real-time properties are analyzed in Section 2.5, and Section 2.6 concludes.

2.2 Reconciliation Framework

This section describes the econometric framework. We start by discussing the measurement equation from our state-space model. Then we explain the structure of the state equation before briefly summarizing the Bayesian estimation approach.

2.2.1 Measurement

Our goal is to reconcile a collection of different output gap estimates (henceforth also referred to as 'base models') in order to obtain the 'true' output gap. To account for the fact that output gaps are often substantially revised over time, we collect several vintages for each point in time and for all methods. Following the notation in Jacobs et al. (2022), we define $y_{t,i}^{t+l}$ to be an estimate of the output gap method $i = 1, \ldots, q$ at time $t = 1, \ldots, T$. The superscript $t + l$ indicates the date on which the estimate became available (vintage or release), with $l = 0, 1, \ldots, n - 1$ denoting the lag of the vintage. For simplicity, we assume that the first estimate is available without lag. Correspondingly, $y_{t,i}^{t}$ is the first estimate (we use the terms 'first', 'initial', and 'real-time' estimate interchangeably), $y_{t,i}^{t+1}$ the second estimate, and so forth, until the last (final) estimate $y_{t,i}^{t+n-1}$. Therefore, we use q different methods and collect n vintages for each time period. Following Jacobs et al. (2022), we assume that each observation can be decomposed according to

$$y_{t,i}^{t+l} = \underbrace{\tilde{y}_t}_{\text{'true'}} + \underbrace{v_{t,i}^{t+l}}_{\text{'news'}} + \underbrace{\xi_{t,i}^{t+l}}_{\text{'noise'}}, \tag{2.1}$$

where $\tilde{y}_t$ is the unobserved 'true' reconciled value of the output gap, $v_{t,i}^{t+l}$ is a news measurement error, and $\xi_{t,i}^{t+l}$ is a noise measurement error.

The 'true' reconciled value of the output gap, $\tilde{y}_t$, is the common component of the underlying estimates of all base models and vintages. It represents the true cyclical state of the economy that is not observed.

Noise measurement errors $\xi_{t,i}^{t+l}$ represent deviations between the reconciled output gap $\tilde{y}_t$ and observations $y_{t,i}^{t+l}$ which are caused by differences in the assumptions, specifications, and parameterizations of the various estimation methods. Since base models rarely agree on a common estimate of the 'true' output gap, noise measurement errors are often persistent and substantial. Therefore, they are correlated with past and future releases, which means that $cov(y_{t,i}^{t+j}, \xi_{t,i}^{t+k}) \neq 0$ for all j and k. However, noise measurement errors are orthogonal to the true values such that $cov(\tilde{y}_t, \xi_{t,i}^{t+l}) = 0$ for all l. When revisions are due to noise measurement errors, current and past vintages of the output gaps contain valuable information that helps predict future revisions.

News measurement errors $v_{t,i}^{t+l}$ are caused by new information that influences the estimate of the 'true' output gap. This new information could be that the GDP estimate for time t is revised or that an estimate of GDP at time $t+1$ becomes available. A revision to the base output gap estimate can thus be caused by a change in the underlying data or a change in the parameters used to estimate the output gap. The news measurement errors are rational forecast errors and correlated with the 'true' output gap, hence $cov(\tilde{y}_t, v_{t,i}^{t+l}) \neq 0$ for all l. The last vintage of a particular observation, observed after a sufficiently long period of time, is then assumed to contain no more news measurement error; hence,

$$y_{t,i}^{t+n-1} = \tilde{y}_t + \xi_{t,i}^{t+n-1}. \tag{2.2}$$

This implies that the last noise measurement error $\xi_{t,i}^{t+n-1}$ captures the remaining deviation of the final vintage from the true output gap.

We define $\mathbf{y}_t$ as a nq-dimensional vector that stacks the different vintages of each model such that $\mathbf{y}_t = [y_{t,1}^{t}, \ldots, y_{t,1}^{t+n-1}, \ldots, y_{t,q}^{t}, \ldots, y_{t,q}^{t+n-1}]'$. Furthermore, we stack the reconciled output gap and both types of measurement errors in a state vector $\boldsymbol{\alpha}_t = [\tilde{y}_t, \mathbf{v}_t, \boldsymbol{\xi}_t]'$. We define m as the number of states equal to $1 + (n-1)q + nq$. The news measurement errors of all models and vintages are contained in $\mathbf{v}_t = [v_{t,1}^t, \ldots, v_{t,1}^{t+n-2}, \ldots, v_{t,q}^t, \ldots, v_{t,q}^{t+n-2}]$ with the exception of the last vintages, which are no longer subject to news shocks. The noise measurement errors are contained in $\boldsymbol{\xi}_t = [\xi_{t,1}^t, \ldots, \xi_{t,1}^{t+n-1}, \ldots, \xi_{t,q}^t, \ldots, \xi_{t,q}^{t+n-1}]$. The measurement equation is then given by

$$\mathbf{y}_t = \mathbf{Z}\boldsymbol{\alpha}_t, \tag{2.3}$$

where $\mathbf{Z}$ is an aggregation matrix of order $nq \times m$, consisting of zeros and ones. An additional error term is omitted. Hence, equation (2.3) encodes the linear additive constraints imposed in equation (2.1). To illustrate, we consider the case of two base models a and b with three vintages for each point in time ($q = 2$, $n = 3$). Equation (2.3) then becomes

$$\begin{bmatrix} y_{t,a}^t \\ y_{t,a}^{t+1} \\ y_{t,a}^{t+2} \\ y_{t,b}^t \\ y_{t,b}^{t+1} \\ y_{t,b}^{t+2} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \tilde{y}_t \\ v_{t,a}^t \\ v_{t,a}^{t+1} \\ v_{t,b}^t \\ v_{t,b}^{t+1} \\ \xi_{t,a}^t \\ \xi_{t,a}^{t+1} \\ \xi_{t,a}^{t+2} \\ \xi_{t,b}^t \\ \xi_{t,b}^{t+1} \\ \xi_{t,b}^{t+2} \end{bmatrix}.$$

This illustration highlights that the last available vintages are not influenced by news errors. Instead, they are only affected by noise measurement errors.

2.2.2 State Process

The state equation follows a vector autoregressive (VAR) process,

$$\boldsymbol{\alpha}_t = \mathbf{T}\boldsymbol{\alpha}_{t-1} + \mathbf{R}\boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \tag{2.4}$$

where the coefficient matrix $\mathbf{T}$ is of order $m \times m$.

We allow both the true output gap and the noise measurement errors to be autocorrelated. Regarding the noise measurement error, it is sensible to assume that there is some persistence over time in the deviation of the output gap estimate from the true output gap. In contrast, news measurement errors do not contain an autoregressive term. As a consequence, $\mathbf{T}$ contains $1 + nq$ nonzero elements on the diagonal:

$$\mathbf{T} = diag(\rho_{\tilde{y}}, \underbrace{0, \ldots, 0}_{\text{News}}, \underbrace{\rho_{\xi_1^0}, \rho_{\xi_1^1}, ..., \rho_{\xi_1^{n-1}}, ..., \rho_{\xi_q^0}, \rho_{\xi_q^1}, ..., \rho_{\xi_q^{n-1}}}_{\text{Noise}}).$$

The m-dimensional vector of the shocks is given by $\boldsymbol{\eta}_t = [\eta_{\tilde{y}_t}, \boldsymbol{\eta}_{v_t}, \boldsymbol{\eta}_{\xi_t}]'$, with news and noise shocks

$$\begin{aligned}
\boldsymbol{\eta}_{v_t} &= \left[\eta_{v_{t,1}^t}, \ldots, \eta_{v_{t,1}^{t+n-2}}, \ldots, \eta_{v_{t,q}^t}, \ldots, \eta_{v_{t,q}^{t+n-2}}\right], \\
\boldsymbol{\eta}_{\xi_t} &= \left[\eta_{\xi_{t,1}^t}, \ldots, \eta_{\xi_{t,1}^{t+n-1}}, \ldots, \eta_{\xi_{t,q}^t}, \ldots, \eta_{\xi_{t,q}^{t+n-1}}\right].
\end{aligned}$$

In the above example, $\mathbf{T}$ and $\boldsymbol{\eta}_t$ become

$$\begin{aligned}
\mathbf{T} &= diag(\rho_{\tilde{y}}, \underbrace{0, \ldots, 0}_{\text{News}}, \underbrace{\rho_{\xi_a^0}, \rho_{\xi_a^1}, \rho_{\xi_a^2}, \rho_{\xi_b^0}, \rho_{\xi_b^1}, \rho_{\xi_b^2}}_{\text{Noise}}), \\
\boldsymbol{\eta}_t &= [\eta_{\tilde{y}_t}, \underbrace{\eta_{v_{t,a}^t}, \eta_{v_{t,a}^{t+1}}, \eta_{v_{t,b}^t}, \eta_{v_{t,b}^{t+1}}}_{\text{News}}, \underbrace{\eta_{\xi_{t,a}^t}, \eta_{\xi_{t,a}^{t+1}}, \eta_{\xi_{t,a}^{t+2}}, \eta_{\xi_{t,b}^t}, \eta_{\xi_{t,b}^{t+1}}, \eta_{\xi_{t,b}^{t+2}}}_{\text{Noise}}]'.
\end{aligned}$$

The structured matrix $\mathbf{R}$ is of order $m \times m$ and determines the co-movement of the shocks $\boldsymbol{\eta}_t$. As a result, $\mathbf{R}$ constrains the behavior of news and noise measurement errors. As shown in Jacobs et al. (2022), the incorporation of multiple data vintages increases the number of observable moments so that the parameters are identified whenever more than one vintage is used. Going back to the basic example with two base models a and b as well as three vintages for each point in time, $\mathbf{R}$ becomes

$$\mathbf{R} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The first $1 + (n-1)q$ elements in the first row are ones. This structure implies that the true output gap is the sum of its lagged value multiplied with $\rho_{\tilde{y}}$, an idiosyncratic shock $\eta_{\tilde{y}_t}$, and $(n-1)q$ news shocks for the output gap at time t. For instance,

$$\tilde{y}_t = \rho_{\tilde{y}} \tilde{y}_{t-1} + \eta_{\tilde{y}_t} + \eta_{v^t_{t,a}} + \eta_{v^{t+1}_{t,a}} + \eta_{v^t_{t,b}} + \eta_{v^{t+1}_{t,b}}. \tag{2.5}$$

The block in $\mathbf{R}$ that corresponds to the news shocks contains values of -1 and 0. This structure implies that the news measurement errors are correlated across vintages. The news measurement error for the output gap at time t, method i, and release l is thus given by the negative sum of current and future news shocks:

$$v^{t+l}_{t,i} = -\sum_{j=l}^{n-2} \eta_{v^{t+j}_{t,i}}, \qquad \text{for } l = 0, \ldots, n-2. \tag{2.6}$$

With three releases, for example, the news measurement error of the first release would consist of the news shock of the first and second releases: $v^t_{t,i} = -\eta_{v^t_{t,i}} - \eta_{v^{t+1}_{t,i}}$. Recall that the last release contains no news measurement error and hence there is also no news shock in the last release. The news measurement error for the penultimate release $n-2$ is then simply the negative value of the news shock of this release, $v_{t,i}^{t+n-2} = -\eta_{v_{t,i}^{t+n-2}}$.

The $m \times m$ covariance matrix $\mathbf{Q}$ determines the co-movement of the shocks $\boldsymbol{\eta}_t$ and is of a block-diagonal form:

$$\mathbf{Q} = \begin{bmatrix} \sigma_{\eta_{\tilde{y}}} & 0 & 0 \\ 0 & \boldsymbol{\sigma}_{\eta_v} & 0 \\ 0 & 0 & \boldsymbol{\sigma}_{\eta_\xi} \end{bmatrix}.$$

This structure implies that the idiosyncratic shock of the true gap is not correlated with any other shock. However, we allow for cross-correlated news shocks in $\boldsymbol{\sigma}_{\eta_v}$ by assuming that news shocks of different base output gap estimates for the same release are correlated, for instance, $cov(\eta_{v_{t,a}^{t+l}}, \eta_{v_{t,b}^{t+l}}) \neq 0$ for any two methods a and b. Recall our example with two base models and three vintages ($\mathbf{v}_t = [v_{t,a}^t, v_{t,a}^{t+1}, v_{t,b}^t, v_{t,b}^{t+1}]'$), the covariance matrix of the news shocks is then of the form

$$\boldsymbol{\sigma}_{\eta_v} = \begin{bmatrix} \sigma_{\eta_{v_a^0}} & 0 & \sigma_{\eta_{v_{a,b}^0}} & 0 \\ 0 & \sigma_{\eta_{v_a^1}} & 0 & \sigma_{\eta_{v_{a,b}^1}} \\ \sigma_{\eta_{v_{a,b}^0}} & 0 & \sigma_{\eta_{v_b^0}} & 0 \\ 0 & \sigma_{\eta_{v_{a,b}^1}} & 0 & \sigma_{\eta_{v_b^1}} \end{bmatrix}.$$

This structure for the news shocks is reasonable, as most output gap methods rely on the same data releases. New information available thus likely shocks each output gap method similarly, for a given release. For flexibility, the covariance matrix of the noise shocks, $\boldsymbol{\sigma}_{\eta_\xi}$, is allowed to be full.

2.2.3 Estimation

Since the binary matrices $\mathbf{Z}$ and $\mathbf{R}$ are fixed and the measurement errors are assumed to be zero, the model parameterization is fairly parsimonious. It requires only the estimation of the state vectors $\boldsymbol{\alpha}_t$, the vector autoregressive coefficients in $\mathbf{T}$, and the covariance matrix of the innovations in the state equation $\mathbf{Q}$.

We use a Normal-Inverse-Wishart prior to estimate $\mathbf{T}$ and $\mathbf{Q}$. Thus, the autoregressive coefficients are assumed to follow a Gaussian distribution, while the covariance matrix of the state equation follows an Inverse Wishart distribution, see Appendix 2.A.1 for further details concerning the prior specification. We use Gibbs sampling to obtain a sequence of draws from the joint posterior distribution. The latent states are estimated using the forward filtering backward sampling algorithm suggested by Carter and Kohn (1994). A set of starting values is randomly generated to ensure robust convergence of the sampler. Furthermore, we assess the convergence of the Gibbs sampling algorithm using trace plots and by checking differences in the recursive means of selected parameters. A detailed description of the estimation algorithm can be found in Appendix 2.A.1.

2.3 Data

We evaluate our methodology for the United States using real-time output gap estimates. In the first subsection, we describe the underlying data (in particular, output). In the second subsection, we explain the different output gap estimation methods which build the database in our model framework.

2.3.1 Real-Time Vintages

Economic data, in particular national accounts, often undergoes substantial revision as new information becomes available. Other reasons for revisions include methodological changes, seasonal adjustment, or temporal disaggregation. Using real-time vintages is, therefore, of crucial importance because they reflect the true information set available to policymakers at a certain point in time. Furthermore, real-time data allow us to take into account the impact of data revisions on the stability of output gap estimates.

We obtain real-time vintages for the United States from Archival Federal Reserve Data (ALFRED), using only the vintages that are available at the release date of the second estimate of GDP. The second estimate is released 55-60 days after the end of the quarter. In contrast to the advance estimate, which is published around one month after the end of a quarter, the second estimate is based on source data that are more complete.[3] We use three time series for the calculation of the different output gap estimates: Real gross domestic product, unemployment rate, and core inflation.[4] We also collect real-time vintages of the Congressional Budget Office's (CBO) potential output estimates. Our quarterly data cover a period from 1960 to the last quarter of 2019 for the United States. However, real-time vintages are only available since 1992.

2.3.2 Base Models

Our model repertoire is based on ten well-established methods. Each method is calculated using five vintages, namely the real-time, second, third, one-year, and two-year vintages ($l = [0, 0.25, 0.5, 1, 2]$). Since most revisions occur within the first two years after the initial publication of GDP (Orphanides and van Norden, 2002; Edge and Rudd, 2016), it is reasonable to not include further releases (thus also limiting the size of the state vector). Five methods are univariate and require only real GDP as input. We apply the popular filter proposed by Hodrick and Prescott (1997), using the parameterization suggested by Ravn and Uhlig (2002). We also use the bandpass filter in Baxter and King (1999) and the algorithm proposed by Hamilton (2018), which takes the two-year forecast error of a projection based on an autoregressive model as the cyclical component of output. Furthermore, we apply an unobserved components model following Watson (1986), which performs a decomposition of GDP into a trend with stochastic drift and a cycle that follows a stationary autoregressive process. For the last and most basic univariate model, we simply fit a cubic spline to logarithmized GDP.

We employ four base models that rely on explanatory variables in addition to the univariate approaches. More precisely, we extend the unobserved component model such that the output gap is modeled as a function of well-observable market outcomes, namely core inflation, unemployment, or both (Kuttner, 1994; Gerlach and

[3] A third estimate is released around one month after the second.
[4] The FRED codes for the three series are GDPC1, UNRATE, and CPILFESL, respectively.

Smets, 1997). The fourth multivariate model is calculated using the Blanchard-Quah decomposition (Blanchard and Quah, 1989; Coibion et al., 2018). They estimate a structural vector autoregression with GDP and the unemployment rate and identify permanent components of GDP growth using long-run restrictions.

Finally, we take real-time vintages of the Congressional Budget Office's potential output estimates to calculate output gaps. The CBO applies the production function approach to measure potential output. Their estimates are widely used in practice and serve as a benchmark in the literature on output gaps (see, for instance, Fernald, 2014; Kamber et al., 2018).

2.4 Results of the Final Estimate

This section evaluates our final estimate, which is calculated as the point-wise median over all draws and estimated using the entire sample ranging to 2019Q4. All results are based on 100,000 draws from the Gibbs sampler, where the first 90,000 are discarded as burn-in. From the remaining 10,000 draws, we keep each fifth draw to further reduce possible autocorrelation between draws.

2.4.1 The Reconciled Output Gap

As a first step, we evaluate whether our reconciled output gap is economically meaningful. Note that the basic requirement of stationarity is met by construction, as it is imposed in the estimation routine. Figure 2.1 compares the reconciled gap with the base output gap models and their median. The reconciled gap is consistent with widely held beliefs about cyclical movements in economic activity, matching the gray-shaded recessions as determined by the National Bureau of Economic Research (NBER). The figure also shows that the 'true' output gap follows the evolution of the underlying base models while tracking the median estimate particularly well. To put this into numbers, Table 2.1 shows summary statistics and, in the last column, the correlation with the reconciled output gap.[5] We find a high correlation of 0.89

[5] A full correlation matrix is available in the Appendix, see Table 2.3.

with the median estimate. Only the multivariate unobserved component estimates (unemployment and trivariate) show a stronger relationship with a correlation of 0.93. The rudimentary cubic spline method is correlated with only 0.43, thus delivering the lowest correlation coefficient from all base models. An important property of our gap is that the mean absolute difference (MD) and the standard deviation (SD) are relatively small across the entire sample for our estimated gap. A smooth estimate is desirable to avoid erratic policy changes (Quast and Wolters, 2022). Also, policymakers may not want to change their policy stance in response to transitory fluctuations, but only to changes in structural forces (St-Amant and van Norden, 1997). The Hamilton filter features the largest measures of dispersion with a standard deviation of 2.4 and mean absolute differences twice the size of most other output gaps.[6] In comparison, our estimate is relatively persistent, indicating that the state-space framework effectively filters out some noise from the often volatile base gap models. While the average of the reconciled output gap is zero, Figure 2.1 also shows that there are shorter periods with a pronounced negative output gap and longer periods with a mildly positive output gap. To quantify this, Table 2.1 shows the minimum and maximum values of the output gaps. For our reconciled value and most other models, the minimum is more negative than the maximum is positive. This result is consistent with the notion of an asymmetric business cycle characterized by relatively substantial negative movements during recessions and smaller amplitudes during expansions (Morley and Piger, 2012).

Finally, Figure 2.10 in the Appendix shows that our reconciled output gap is reasonably precise. The 95% posterior mass is tightly centered around the pointwise median and rarely covers the zero line. This is important from a policymaker's perspective, as the explanatory power of an output gap diminishes with statistical uncertainty. The credible interval increases slightly towards the end of the sample, reflecting higher uncertainty when only the first few releases are available.

[6] A reason for this volatile behavior of the Hamilton filter, as shown by Quast and Wolters (2022), is that it passes through and amplifies some high-frequency noise present in the data.

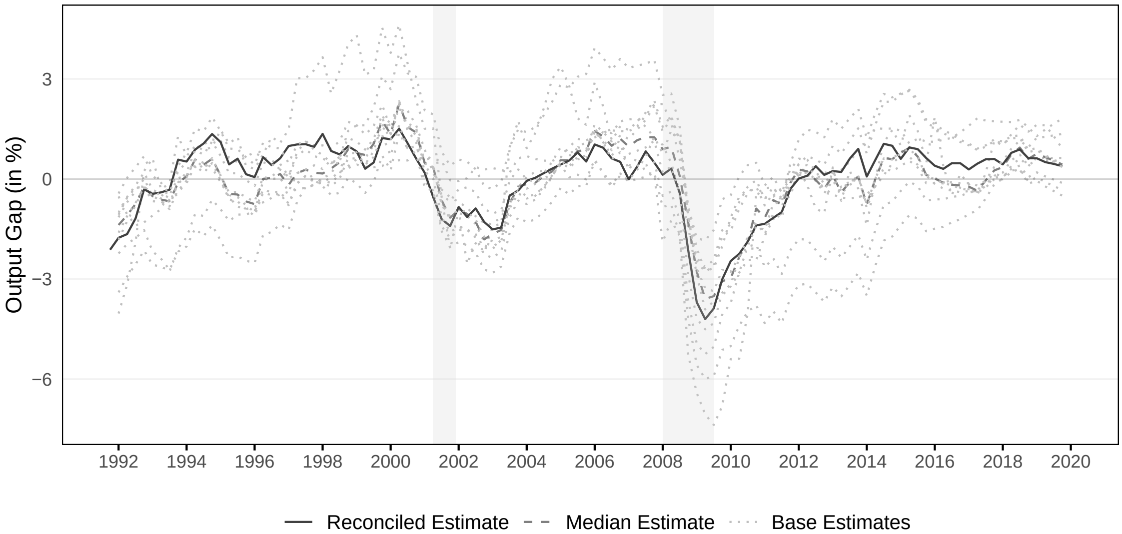


Figure 2.1: Final output gap estimates. Notes: This figure shows the reconciled, median, and the ten base output gap estimates in quarterly frequency from 1992 to 2019. All output gaps are estimated using the full (final) sample. Shaded areas highlight recessions as determined by the NBER Business Cycle Dating Committee.

Output Gap Method	Mean	Min	Max	SD	MD	Cor
Reconciled Estimate	0.00	-4.20	1.51	1.14	0.32	-
Median Estimate	-0.06	-3.60	2.25	1.07	0.33	0.89
Baxter-King Filter	0.04	-2.75	2.25	0.95	0.24	0.68
Blanchard-Quah Decomposition	0.35	-3.92	1.82	1.25	0.22	0.83
Congressional Budget Office	-1.28	-6.00	2.31	1.87	0.38	0.78
Hamilton Filter	0.34	-7.38	4.63	2.42	0.75	0.88
Hodrick-Prescott Filter	-0.03	-2.79	2.35	1.01	0.37	0.72
Spline	-0.05	-3.37	3.93	2.04	0.38	0.43
Unobserved Components	-0.02	-2.64	2.22	0.92	0.36	0.71
Unobs. Comp. (Inflation)	-0.01	-1.84	1.46	0.55	0.34	0.57
Unobs. Comp. (Unemployment)	0.21	-5.25	2.64	1.59	0.38	0.93
Unobs. Comp. (Trivariate)	0.15	-4.51	2.68	1.45	0.38	0.93

Table 2.1. Summary statistics for the reconciled and base output gaps. Notes: This table shows summary statistics for all output gap methods. Min and Max refer to the minimum and the maximum, respectively. SD denotes the standard deviation and MD the mean absolute difference. The last column (Cor) shows the correlation of an output gap with the reconciled estimate. All measures are calculated based on the full sample estimates.

2.4.2 News and Noise Measurement Errors

Figure 2.2 shows the news measurement errors of the different releases and methods over time. We see that the real-time releases (black lines) contain larger news components in all cases. This makes intuitive sense as the amount of news arguably decreases with each release, thus supporting our choice to not include vintages older than two years.

The news measurement errors of the first release are particularly large for the CBO estimate and the Baxter-King filter. This implies that the total revision between the final and the initial estimate incorporates a substantial news component, which is informative in calculating the reconciled output gap.[7] Note that large revisions are not in general accompanied by sizable news components. An example is the Spline method, which, as discussed in the next section, often undergoes considerable revision. Given the relatively small news measurement errors, we can conclude that the Spline revisions are largely driven by noise. We also observe that the revisions from the Hodrick-Prescott filter and the unobserved component models feature small news components. Considering the second, third, and one-year-later vintages, the magnitude of the news errors decreases only slightly across releases for the Blanchard-Quah decomposition. In other words, later releases have an equally important role as information that is published earlier.

Figure 2.11 in the Appendix shows the noise measurement errors. In contrast to the news errors, the noise errors are large and show substantial persistence over time. This implies that the output gaps of the base models generally deviate persistently from the reconciled gap, thus reflecting the different methodologies involved in estimating the output gap. In real-time, the Blanchard-Quah decomposition and the HP filter exhibit the lowest noise errors. Two years after the real-time release, the unobserved components model and the HP filter contain only a small amount of

[7] This is an implication of equation 2.1 and 2.2, as the total revision can be decomposed according to the formula $y_{t,i}^{t+n-1} - y_{t,i}^{t} = -v_{t,i}^{t} + \text{noise}$, where $v_{t,i}^{t}$ is the news error from the real-time release.

noise, which suggests that they are the closest to our reconciled measure in absolute terms.[8]

2.5 Results of the Real-Time Estimates

Although the reconciled output gap shows desirable full-sample characteristics, this does not need to hold in the case of real-time data. In this section, we evaluate the real-time reliability of the reconciled and base output gaps in the spirit of Orphanides and van Norden (2002). More specifically, we calculate each gap using real-time data vintages and then compare them to those based on revised data.

We show all vintages over time in Section 2.5.1. In Section 2.5.2 we evaluate the revisions of all methods and plot the revisions over lags and time. In Section 2.5.3, we compare the symmetry of the output gaps. Subsequently, we evaluate the forecasting ability of the gaps in Section 2.5.4 for inflation. Finally, Section 2.5.5 discusses the results from the Kalman gain, which shows how much each method contributes to the reconciled gap.

2.5.1 Vintages

Figure 2.3 shows the vintages of the output gaps in quarterly frequency, with the first vintage from 2000Q1 and the last vintage from 2019Q4. Earlier vintages are plotted in light gray, while the latest release is a black line. Note that we used the last vintage to compute the final estimate of the reconciled output gap in the previous section. The figure highlights two important observations. First, the resulting model estimates differ remarkably in terms of the size of the revisions. The output gap based most on expert knowledge, i.e., the method by the CBO, appears to be subject to strong revisions over time. Although in amplitude similar to the CBO estimate, the Hamilton filter is more reliable in terms of real-time estimation. One reason for its stability is that data revisions only lead to small changes in the parameters of the

[8] Recall equation 2.2 which implies that the deviation of the last release from the reconciled gap is the noise error of the last release.

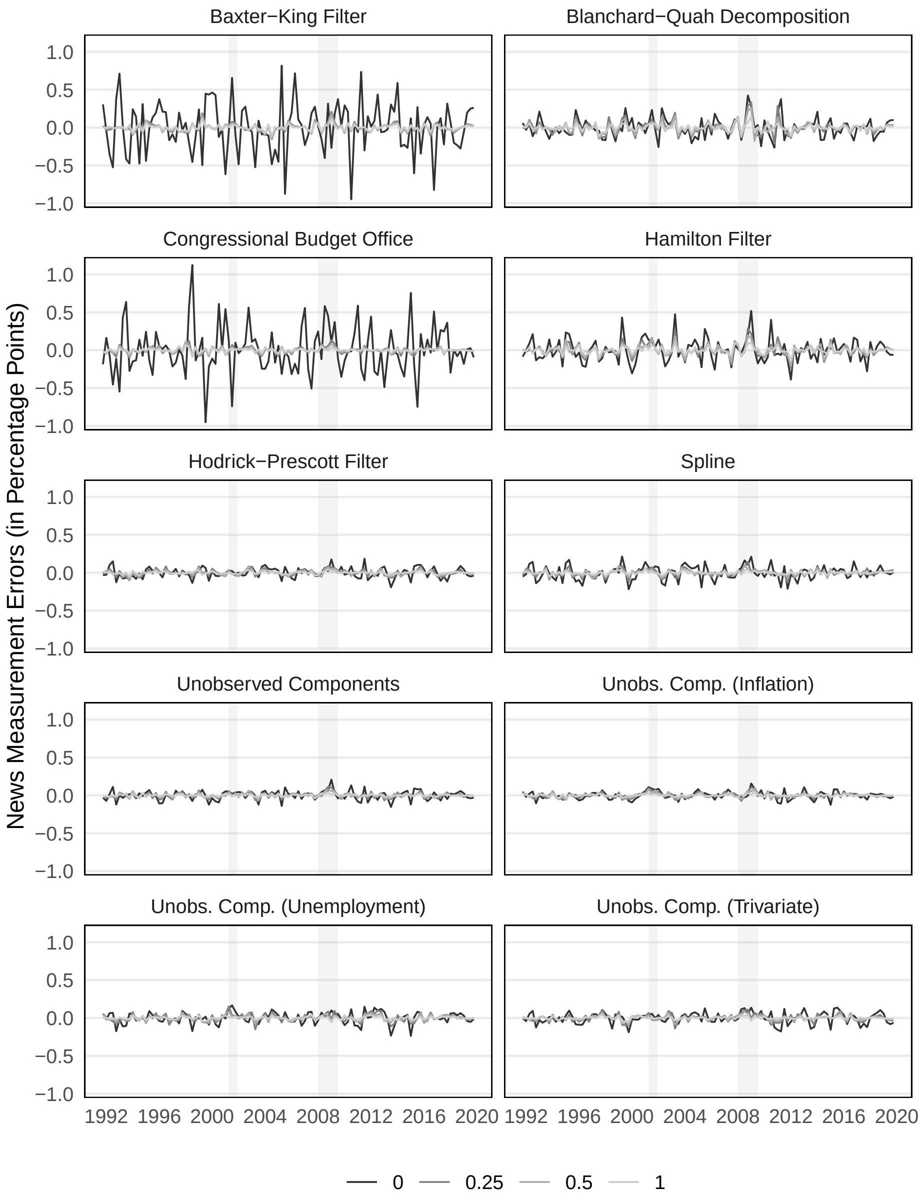


Figure 2.2: News measurement errors. Notes: This figure shows the estimated news measurement errors for each output gap method using the full sample. The black lines depict the news errors from the real-time release, while the dark, middle, and light gray lines do so for the second release, the third release, and the release that is published one year after the real-time release, respectively. Shaded areas highlight recessions as determined by the NBER Business Cycle Dating Committee.

univariate AR model (Hamilton, 2018; Quast and Wolters, 2022). Another reason is the Hamilton filter's one-sided nature, thus avoiding large end-point problems present in two-sided filters such as the HP filter (Kamber et al., 2018). Although sharing a similar methodology, we find pronounced differences between the included unobserved component models: The univariate model and the one including inflation show a distinct tendency to revise around the financial crisis, while the model estimate that includes unemployment is one of the models least affected by future releases. The median estimate also undergoes adjustment around the financial crisis, while the reconciled estimate is fairly stable over the entire horizon.

Second, the figure highlights the discrepancies between the output gap estimates. Substantial differences exist between the various methods in terms of magnitude and persistence, even several years after the initial release. During the early 2000s recession, for example, the final estimates for 2001Q1 range from -0.3% for the unobserved components model with unemployment to 2.1% for the Spline. Since we assume that all methods attempt to track the true state of the output gap, these differences between the final estimates can only be attributed to noise measurement errors. As a consequence of the diverse set of methods involved, these errors are large and persist over time (see Figure 2.11).

2.5.2 Revisions

Since potential output and the output gap are not observable, there is no way to quantify the precision of any given estimation method. Therefore, a common approach is to evaluate certain properties that are important for the intended application. An important criterion in our application is real-time reliability, as the initial output gap estimate is often revised substantially when new information becomes available (Orphanides and van Norden, 2002). This criterion is crucial for policymakers who require a timely and accurate assessment of the cyclical position of an economy. It is thus natural to evaluate our reconciled output gap and also all other base gaps by comparing the revisions over lags and time in a real-time study. Revisions are calculated using root mean squared differences to the final estimate. Furthermore, we use a nonparametric Friedman rank test to check for

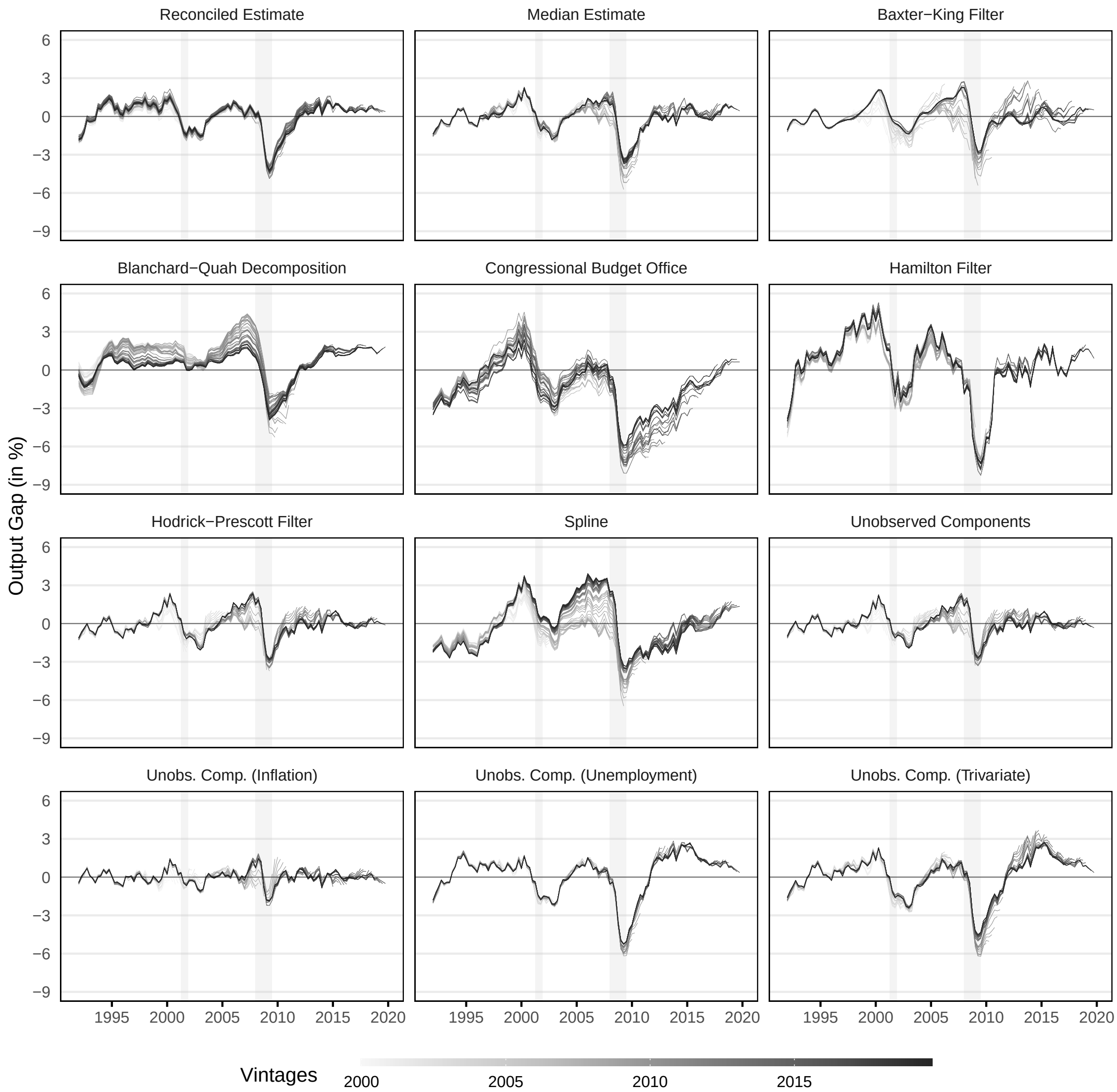


Figure 2.3: Real-time vintages. Notes: This figure shows real-time vintages of output gap estimates in quarterly frequency. The first vintage is from 2000Q1 and the last vintage from 2019Q4. Shaded areas highlight recessions as determined by the NBER Business Cycle Dating Committee.

significant differences between the revisions of the different models (Kourentzes and Athanasopoulos, 2019).

Figure 2.4 shows the revisions, measured as the root mean squared changes between the h-th estimate and the final estimates for the same period where h

ranges from zero to four years. For comparing real-time estimates and revised data, we discard the last two years of observations in order to prevent bias from converging estimates (following Orphanides and van Norden, 2002; Quast and Wolters, 2022). Hence, our real-time evaluation of output gaps ends with the estimate from 2017Q4 whereas the final estimate is still the vintage from 2019Q4. Figure 2.4 provides evidence that our reconciliation framework leads to more resilient estimates than the underlying base models. In particular, the root mean squared difference between the initial (year 0) and the final estimate is the lowest for our reconciled gap. Note that the revisions of our reconciled estimate are also smaller than those of the median estimate, although our estimate converges to the median model after two years. Confirming the earlier graphical observation, a base gap with favorable revision properties is the unobserved component model that includes unemployment (dotted line positioned slightly above our gap in real-time), while the dotted line at the top (i.e., the largest revisions) represents the cubic spline. The revision pattern of the Blanchard-Quah decomposition is unusual since revisions tend to increase across lags,[9] which suggests that this method suffers from parameter instability.

But is our reconciled output gap only stable on average? In other words, is our estimate prone to revisions during, for example, the financial crisis? Figure 2.5 shows the average revision for each year, given by the root mean squared changes between the initial and the final output gap estimates. It provides evidence that revisions for the base models as well as the combined estimates are generally higher during periods of pronounced economic volatility. This can be attributed to model instabilities and larger revisions of the underlying statistics. Both the reconciled and the median model are often more resilient to revisions than the underlying base models. These findings are consistent with the literature on forecast combination (see, for instance, Timmermann, 2006; Pettenuzzo and Timmermann, 2017). They show that model pooling does not necessarily outperform the best model, but leads to more precise predictions on average. In contrast to the median estimate, however,

[9]For example, for 2009Q1, the real-time estimate is -3.6%, after one year -2.4%, after three years -3.2%, and the final estimate -2.5%. The absolute differences to the final estimate are 1.1, 0.1, and 0.7, thus increasing again after the one-year release.

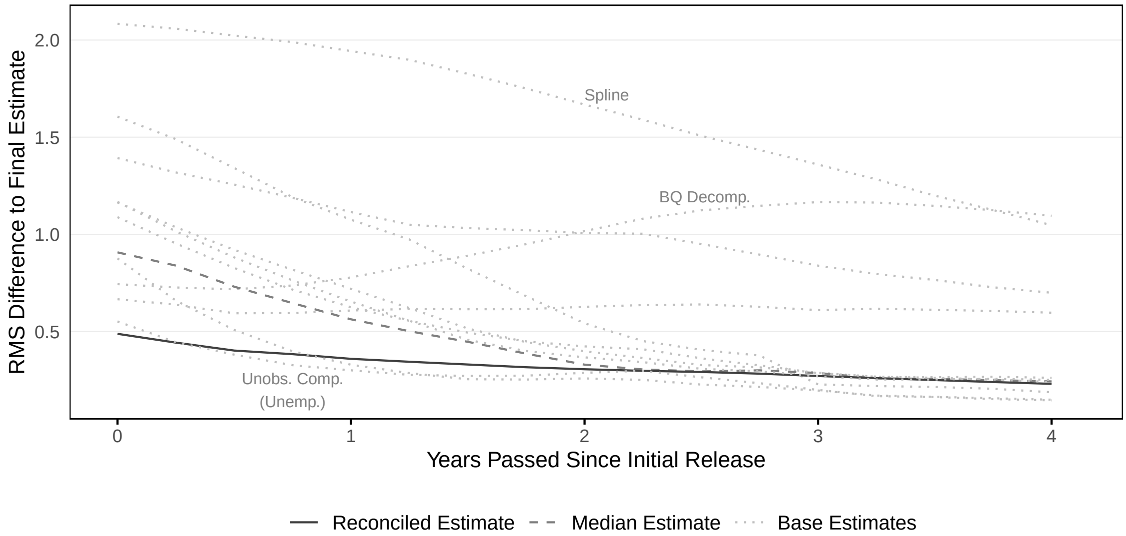


Figure 2.4: Root mean squared difference between the h-th and the final estimate. Notes: This figure shows revisions as the root mean squared changes between the h-th and the final estimate where h ranges from zero to four years. The reconciled, median, and the ten base estimates are shown. We consider revisions until 2017Q4 and use the 2019Q4 vintage as the final estimate.

our reconciled gap performs among the best methods during both the recession in the early 2000s and the financial crisis.

So far, we have focused only on a graphical treatment of unadjusted root mean squared revisions. Following the literature on output gaps, we now evaluate alternative revision measures for each individual method. Orphanides and van Norden (2002) proposed several measures to quantitatively assess the relative importance of revisions to an output gap series (subsequently also used in, for instance, Fleischman and Roberts, 2011; Kamber et al., 2018; Barbarino et al., 2020; Quast and Wolters, 2022). In Table 2.2, we show these revision and reliability indicators.[10] The second and third columns show the correlation of the initial with the final estimate, and the frequency of sign changes between the two estimates, respectively. Estimating the correct sign in real-time is important, as a positive output gap suggests measures

[10]Since our analysis covers a different time period than that of Orphanides and van Norden (2002), the statistics are not directly comparable.

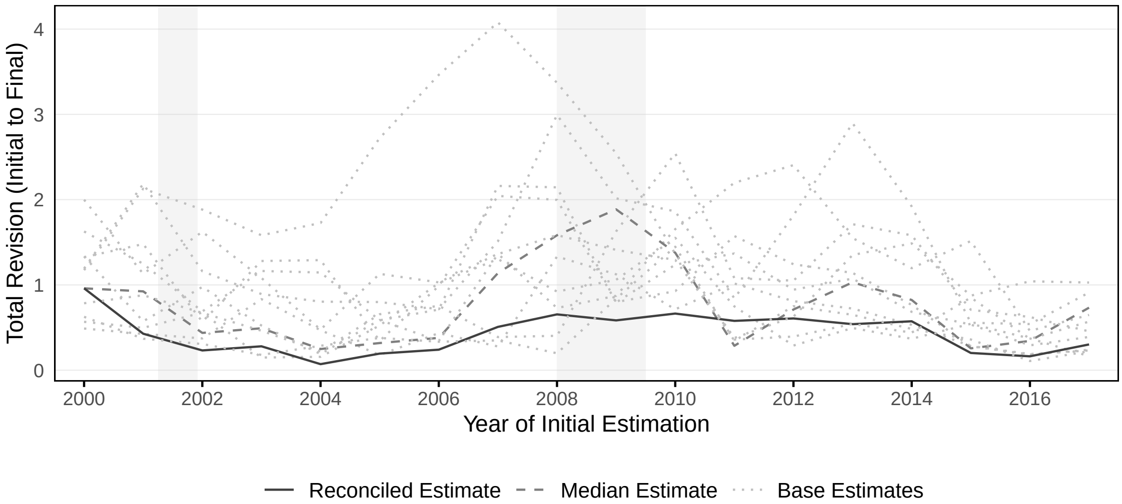


Figure 2.5: Revisions by estimation date. Notes: This figure shows the average annual revisions as the root mean squared changes between the initial and the final estimate for the reconciled, median, and ten base estimates. Shaded areas highlight recessions as determined by the NBER Business Cycle Dating Committee. We consider revisions until 2017Q4 and use the 2019Q4 vintage as the final estimate.

to slow down the economy, while a negative gap may call for fiscal or monetary stimulus. We find almost one-third of initial estimates from the median model to flip sign. The reconciled estimate does so in 10% of the cases. The unobserved component models that include unemployment (i.e., also the trivariate model) are especially robust to sign changes (3% and 4%, respectively) while including only inflation degrades the performance to a coin toss. The fifth and sixth columns show the standard deviation and root mean squared differences of the revisions.[11] In order to adjust for the method-specific variance of both the final estimate and the revisions, we calculate the ratio of the revisions' standard deviation (NS) and root mean squared difference (NSR) to the standard deviation of the final estimate (last two columns). Both measures are proxies for the noise-to-signal ratio in the real-time estimates. An NSR of one would indicate, for instance, that the ex-post revisions are on average of the same size as one standard deviation of the ex-post estimate

[11]The information contained in the RMS column corresponds to the position of each line at horizon zero in Figure 2.4.

of the cycle (Orphanides and van Norden, 2002). For the median gap, the noise-to-signal ratios are almost twice as high as for the reconciled output gap. Only the Hamilton filter and the unobserved component model that includes unemployment have a lower ratio in our sample. We thus confirm the results by Fleischman and Roberts (2011) for the US and Furlanetto et al. (2023) for Norway, as they find output gap models that include unemployment to be reliable estimates of the cyclical state of the economy. We also find that the Baxter-King filter has the highest NSR among all methods. This is arguably a result of its symmetric filter nature, thus being subject to pronounced end-of-sample instability (Mise et al., 2005).

	Initial vs. Final		Final	Revisions		Ratios	
Output Gap Method	COR	SIGN	SD	SD	RMS	NS	NSR
Reconciled Estimate	0.93	0.10	1.24	0.49	0.49	0.39	0.39
Median Estimate	0.86	0.32	1.24	0.90	0.91	0.72	0.73
Baxter-King Filter	0.53	0.38	1.10	1.60	1.61	1.46	1.46
Blanchard-Quah Decomp.	0.98	0.08	1.42	0.71	0.74	0.50	0.52
Congressional Budget Office	0.89	0.12	1.84	1.29	1.39	0.70	0.76
Hamilton Filter	0.98	0.15	2.56	0.55	0.67	0.22	0.26
Hodrick-Prescott Filter	0.49	0.42	1.18	1.17	1.16	1.00	0.99
Spline	0.62	0.42	2.20	1.78	2.08	0.81	0.95
Unobserved Components	0.37	0.44	1.08	1.17	1.17	1.09	1.09
Unobs. Comp. (Inflation)	0.04	0.50	0.63	0.88	0.88	1.40	1.39
Unobs. Comp. (Unemp.)	0.98	0.03	1.88	0.55	0.55	0.29	0.29
Unobs. Comp. (Trivariate)	0.95	0.04	1.71	1.10	1.09	0.64	0.64

Table 2.2. Summary revision and reliability indicators. Notes: The table shows measures evaluating the size, sign, and variation for all considered methods. COR denotes the correlation between real-time and final estimates. SIGN denotes the frequency with which the real-time and final gap estimates have opposite signs. Final SD gives the standard deviation of the final estimate and Revisions SD of the revisions (the difference between the final and the real-time value). RMS denotes the root mean squared revisions. NS denotes the ratio of the standard deviation of the revision to that of the final estimate of the gap. NSR denotes the ratio of the root mean square of the revision to the standard deviation of the final estimate of the output gap. We consider vintages until 2017Q4 and use the 2019Q4 vintage as the final estimate.

In order to assess the statistical significance of these findings, it is also instructive to rank the methods according to the noise-to-signal ratio for each date and lag.

Following Kourentzes and Athanasopoulos (2019), we use non-parametric Friedman and post-hoc Nemenyi tests to check for significant differences in the ranks of the NSRs (Hollander et al., 2015). More precisely, for each lag from 0 to 4 years (17 horizons) and dates 2000Q1 to 2017Q4, we calculate the root squared revisions divided by the standard deviation of the final gap and rank the 12 methods by this noise-to-signal ratio. We then calculate the mean of these ranks across all dates and lags. The resulting mean ranks for each method are shown in Figure 2.6, where a rank of one would indicate that a method has the smallest NSR for each lag and date. In addition, we apply a post-hoc Nemenyi test to determine whether there are groups of models that perform similarly. Based on the calculated critical distance of 0.48, the shaded areas indicate groups of models that are not significantly different from each other. We find two clusters of models that rank similarly in terms of NSR and one outlier on both sides of the spectrum. The unobserved component model that includes unemployment features a mean rank of close to three, while the spline is located between 9 and 10. The reconciled output gap has the second lowest mean rank of 5. According to the post-hoc Nemenyi test, the reconciled gap shares the noise-to-signal ratio with the Hamilton filter and the trivariate unobserved component model. The median estimate is also allocated in the first cluster but has a significantly higher mean rank than the reconciled estimate. The output gap that is often taken as a reference series, i.e., the CBO estimate, ranks poorly and shares a place with the Blanchard-Quah decomposition in the last cluster.

2.5.3 Output Gap Symmetry

Fiscal policymakers rely on output gap estimates to perform cyclical adjustments of fiscal balances. To determine a structurally balanced budget, it is common to estimate levels of spending that are independent of the business cycle. Mechanically, this leads to a counter-cyclical fiscal policy since revenues are, in general, strongly linked to the business cycle (Vegh and Vuletin, 2015). To be useful in such a context, a trend-cycle decomposition should produce a symmetric output gap and, in turn, balanced government spending. Since fiscal planning depends on estimates at the current edge, the real-time properties of output gap measures are again of crucial

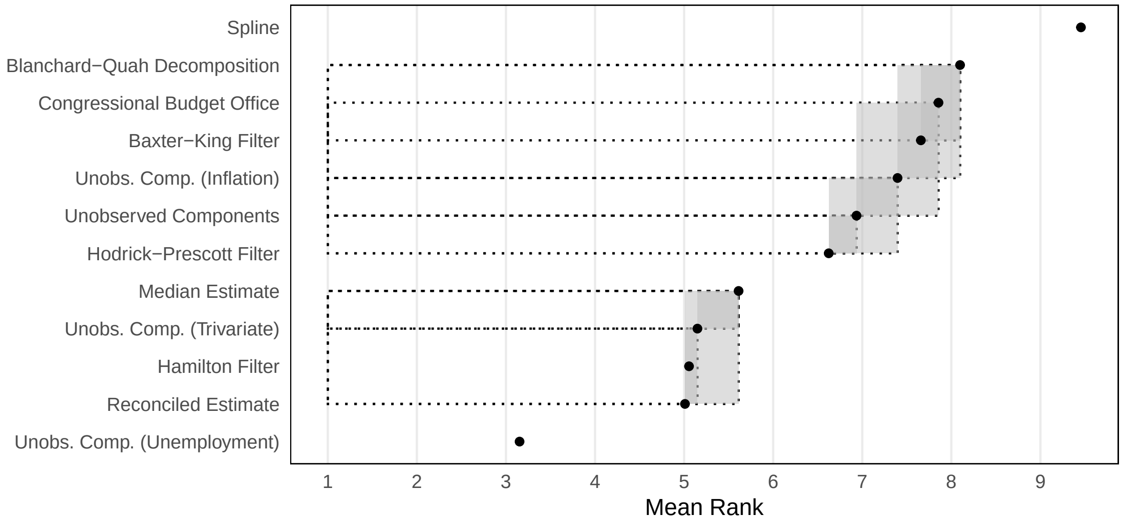


Figure 2.6: Significance test of ranked noise-to-signal ratio. Notes: This figure shows the results from a non-parametric Friedman rank test. Mean ranks indicate the average order when ranking the noise-to-signal ratios (root squared revisions divided by the standard deviation of the final gap) for each date and lag. Shaded areas indicate groups of models that are not different from each other at the 95% confidence level, based on post-hoc Nemenyi tests. We consider revisions until 2017Q4 and use the 2019Q4 vintage as the final estimate.

importance. Figure 2.7 shows the average output gap for each method for the series of initial estimates as well as for the final vintage. We find most output gaps to be fairly symmetric both in real-time and ex-post. Between 2000 and 2019, the average of our model's real-time estimates is slightly negative at -0.2%. In our view, this result is reasonable given that the Great Recession took place in this sample span. The CBO exhibits a substantial downward bias both in real-time (-2.1%) and over the course of the final vintage (-1.3%). In contrast, the Blanchard-Quah decomposition and the Hamilton filter suggest that the economy has run above potential, on average, in real-time and from a retrospective point.

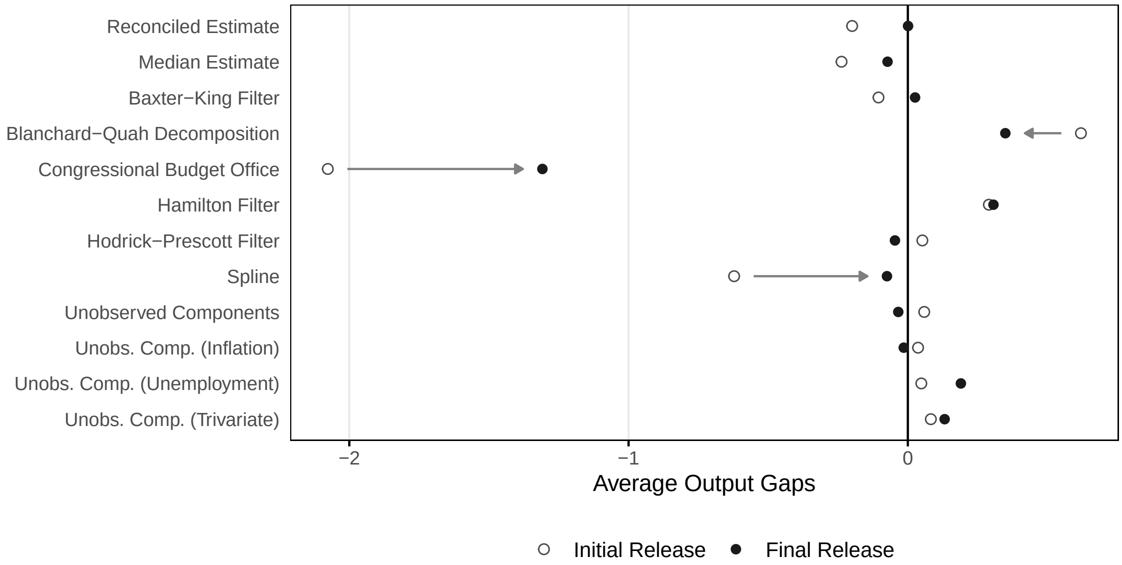


Figure 2.7: Output gap symmetry. Notes: This figure shows the average output gap, based on initial and final releases (i.e., based on the 2019Q4 vintage). Arrows indicate real-time biases.

2.5.4 Inflation Forecasts

As the output gap is used by monetary policy authorities to assess the inflationary pressure in an economy, an output gap estimate is more informative if it can forecast inflation. To test the real-time ability of an output gap to predict inflation, we perform an out-of-sample forecasting exercise with the following autoregressive distributed lag (ADL) Phillips curve forecasting equation (Kamber et al., 2018; Quast and Wolters, 2022):

$$\pi_{t+h} - \pi_t = \alpha + \sum_{p=1}^{n_\pi} \beta_p \Delta\pi_{t-p} + \sum_{p=1}^{n_y} \gamma_p y_{t-p,i}^{t+h} + \epsilon_{t+h}, \qquad \text{for } t = 1, \dots, T-h. \tag{2.7}$$

where π_t denotes the quarterly rate of US inflation at time t. As explanatory variables, we include the difference of the quarter-on-quarter inflation rate with lags 1 to n_π (or no lag of the inflation rate) and the real-time output gap of method i with lags 1 to n_y. Equivalent to the notation in section 2.2, $y_{t,i}^{t+h}$ denotes the output gap at time t of method i estimated with data available until time $t+h$ (since the inflation rate of time $t+h$ is available). We estimate the equations by ordinary least squares for all methods, vintages 2000 to 2019, and forecast horizon h from one to four quarters. We include up to four lags of the inflation rate and the output gap and choose the best model specification based on the Bayesian Information Criterion (BIC). In most regressions, the BIC is the lowest if both variables are included with only one lag. The coefficients for the first lagged output gap are positive in almost all cases, as expected. We use the estimated coefficients, $\hat{\beta}_p$ and $\hat{\gamma}_p$, to predict inflation with forecast horizon h and calculate forecast errors as the difference between the predicted and the observed inflation rates.

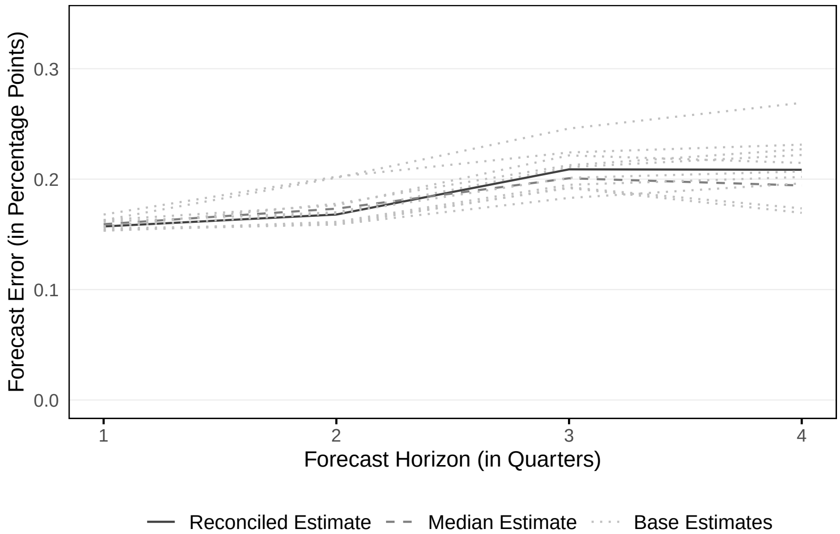


Figure 2.8: Inflation forecast error. Notes: This figure shows root mean squared forecast errors from pseudo out-of-sample predictions of the inflation rate from different output gap estimation methods at various forecast horizons (one to four quarters). The output gap data starts in 2000Q1 and ends in 2018Q4, the inflation rates are from 2019Q4.

Figure 2.8 shows the root mean squared forecast error per horizon and method. Our reconciled estimate is among the methods with the lowest forecast errors for the first two quarters while lying above the median model for the third and fourth quarters. In order to test whether some of the computed forecasts possess significant predictive superiority, we apply the modified Diebold-Mariano (DM) test (Diebold and Mariano, 1995; Harvey et al., 1997; Guérin et al., 2015). With eleven methods (ten base gaps and the median estimate) and four forecasting horizons, we conduct the modified DM test 44 times, each time comparing the loss differential between the forecasts from the reconciled estimate and the ones from an alternative method. At the 5% significance level, we cannot reject the null hypothesis of equal predictive accuracy in 40 cases.[12] At the 1% significance level, only the forecast from the Hamilton filter is better for the third quarter, while we fail to reject the null for all other 43 tests. This exercise shows that in our sample no single output gap method forecasts inflation significantly better for all horizons. In addition, forecasting inflation only with the past inflation rate as an explanatory variable leads to similar forecast errors.[13] This result is in line with the literature, as univariate inflation forecasting models often perform similarly or even better than models which use output gaps as an explanatory variable, especially in recent decades (Stock and Watson, 2007; Dotsey et al., 2018; Kamber et al., 2018; Quast and Wolters, 2022).

2.5.5 Relative Contributions of the Output Gap Estimation Methods

Following Jacobs et al. (2022), we use the estimated Kalman gains to assess the importance of the different base models and different releases with respect to the reconciled output gap. Figure 2.9 shows the average weights that the reconciled gap puts on each release and base model at a given lag. In real-time (lag 0), the Kalman filter is by construction only able to put weight on the real-time release, as later releases are not yet available. Although the reconciled output gap features reliable real-time characteristics, we find that this is only possible thanks to the inclusion of noisy estimates like the Spline or Blanchard-Quah decomposition, which

[12] The Hamilton filter provides a significantly better forecast for one forecast horizon (3 quarters) and the Hodrick-Prescott filter also for one horizon (4 quarters). The Blanchard-Quah decomposition yields a significantly worse forecast than our reconciled estimate for two forecast horizons (2 and 3 quarters).

[13] According to the modified Diebold-Mariano test, a forecasting model containing only the inflation rate performs better in two out of 48 tests and worse in four out of 48 tests. In the remaining 42 comparisons of the forecasting accuracy, the differences are not statistically significant.

feature a Kalman gain close to 0.2 in real-time. Moving forward, we also find that the algorithm puts the largest weight on the newest release in most cases, thereby pricing in the most recent information available. This does not mean, however, that the inclusion of multiple vintages is pointless as earlier releases still carry weight in computing the 'true' output gap. After two years, the releases typically share relatively similar filter gains. We also observe that the relative importance of each method converges over time. The Baxter-King filter, for example, is only gaining weight after the second and third release, while the importance of the Spline decreases. In summary, our reconciled estimate is a combined estimate in the sense that it exploits information from all methods and releases.

2.6 Conclusion

We use a state-space framework to estimate a 'true' output gap that is reconciled from ten well-established output gap methods. The inclusion of multiple releases helps with optimally extracting information from noisy output gap estimates, identifying news and noise components in vintages, and modeling revisions for each method. We show that this leads to an intuitive and economically meaningful output gap estimate. In a comprehensive real-time study for the United States, we review the performance of all output gap methods and compare them to a median model and our reconciled gap. We provide evidence that our output gap features the lowest root mean squared revisions in real-time over the entire sample-span, while also being reliable in times of crisis. In contrast to benchmark methods like the Congressional Budget Office estimate, the reconciled gap is characterized by symmetry both in real-time and ex-post, thus advocating its use for balanced fiscal planning. In summary, our results point to the importance of incorporating multiple releases and methods to build a reliable output gap.

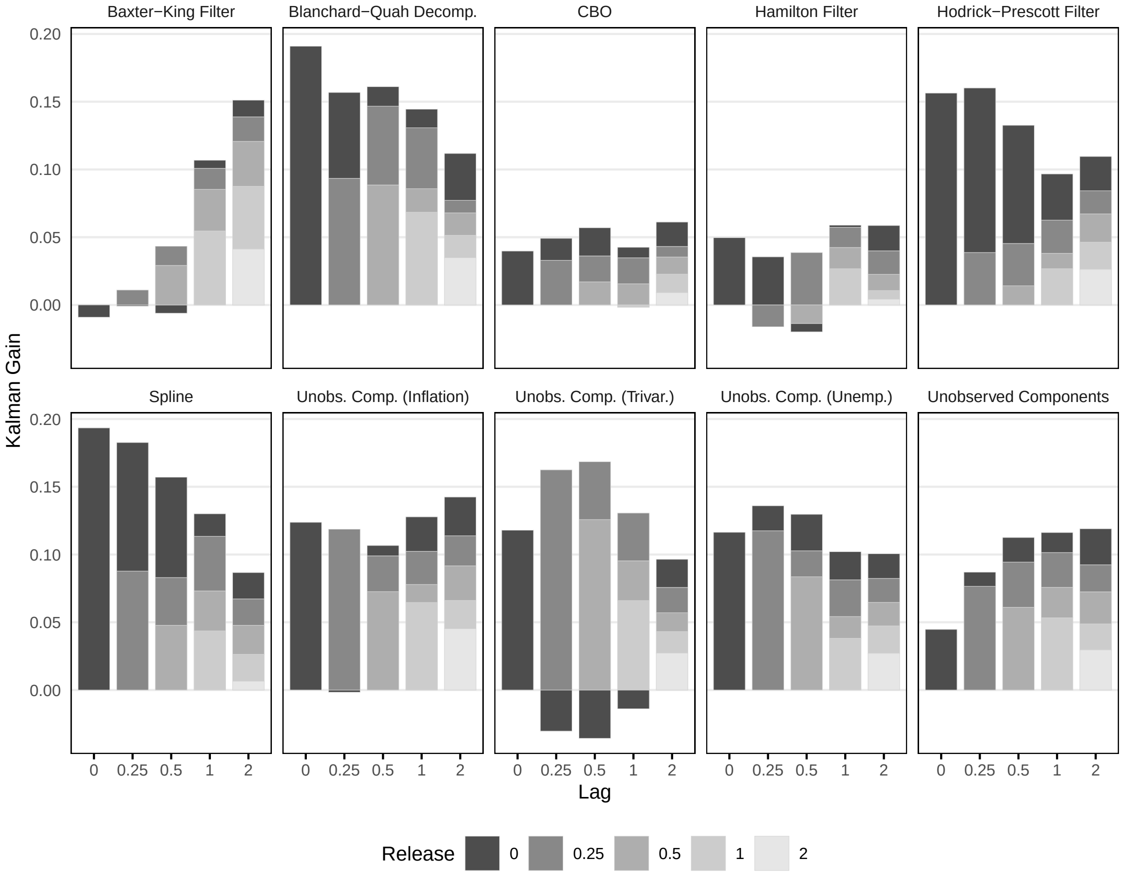


Figure 2.9: Kalman gains. Notes: This figure shows for each output gap the average Kalman gains assigned to each release (grey blocks) at a given lag. The lag is defined as the difference between vintage and date. The first bars (lag=0) show the Kalman gains when only the first release is available (i.e., in real-time). The stacked second bars (lag=0.25) show the gains when the first two releases (0 and 0.25) are available. The last bars (lag=2) show the Kalman gains when all five releases are available. The figure can be read, for instance, that one year after the initial release (lag=1) the unobserved components model with unemployment receives a weight of approximately 0.1, where the release with a lag of one year (second-brightest block, newest release available at that time) receives the highest weight of all releases (0.04). However, also the other releases (0, 0.25, and 0.5) have a non-negligible weight.

2.A Appendix

2.A.1 Sampling Algorithm

The dynamic factor model benefits from a parsimonious parameterization. The factor loadings are given by a fixed aggregation matrix $\mathbf{Z}$, consisting of ones and zeros. Similarly, the structured matrix $\mathbf{R}$ is known ex-ante and requires no estimation. The matrix $\mathbf{T}$ determines the state process and consists of a $1 + nq$ non-zero element on the diagonal. The autoregressive coefficients are assumed to follow a normal distribution. Each nonzero coefficient is shrunk towards zero a priori, i.e., $\rho_i \sim N(0, \frac{1}{16})$. Note that we impose stationarity for each draw.

The covariance matrix of the state equation $\mathbf{Q}$ is sampled from an inverse Wishart distribution. The prior follows $\mathbf{Q} \sim IW(Q_0, T_0)$ where Q_0 is an $m \times m$ scale matrix with diagonal entries $[0.8, 0.2, \ldots, 0.2]$ and $T_0 = m + 1$ is the prior shape. The latent state vectors $\boldsymbol{\alpha}_t$, containing the true output gap as well as the news and noise measurement errors, are demeaned after being estimated by forward-filtering backward-sampling (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994).

Our data is demeaned prior to estimation and includes $q = 10$ output gap methods and $n = 5$ releases. We consider the following lags: $l = [0, 0.25, 0.5, 1, 2]$ where a difference of 0.25 indicates a quarterly step. Hence, we assume that the estimated output gap is final two years after the initial estimate. To reduce the size of the state vector, we abstract from including all quarterly estimates.

The model parameters are estimated using Gibbs sampling. We discard 90,000 draws as burn-in and cycle through an additional 10,000 iterations, saving every fifth draw such that our posterior analysis is based on 2000 draws. The convergence of the sampler is ensured through visual inspection of recursive mean plots and convergence diagnostics on selected chains (Geweke, 1991).

2.A.2 Tables and Figures

Output Gap Method	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) Reconciled Estimate	1.00	-	-	-	-	-	-	-	-	-	-	-
(2) Median Estimate	0.89	1.00	-	-	-	-	-	-	-	-	-	-
(3) Baxter-King Filter	0.68	0.90	1.00	-	-	-	-	-	-	-	-	-
(4) Blanchard-Quah Decomp.	0.83	0.78	0.60	1.00	-	-	-	-	-	-	-	-
(5) Congressional Budget Office	0.78	0.82	0.71	0.69	1.00	-	-	-	-	-	-	-
(6) Hamilton Filter	0.88	0.85	0.64	0.68	0.84	1.00	-	-	-	-	-	-
(7) Hodrick-Prescott Filter	0.72	0.91	0.95	0.58	0.68	0.65	1.00	-	-	-	-	-
(8) Spline	0.43	0.69	0.79	0.54	0.73	0.54	0.72	1.00	-	-	-	-
(9) Unobserved Components	0.71	0.90	0.94	0.54	0.64	0.63	0.99	0.66	1.00	-	-	-
(10) Unobs. Comp. (Inflation)	0.57	0.70	0.74	0.35	0.43	0.48	0.82	0.44	0.86	1.00	-	-
(11) Unobs. Comp. (Unemp.)	0.93	0.81	0.56	0.80	0.59	0.78	0.64	0.25	0.65	0.54	1.00	-
(12) Unobs. Comp. (Trivariate)	0.93	0.88	0.69	0.81	0.68	0.78	0.77	0.41	0.77	0.61	0.96	1.00

Table 2.3. Correlation table of the reconciled and base output gap estimates. This table shows the correlations for each output gap method. All measures are calculated based on the full sample estimates.

Figure 2.10: Final output gap estimates with 95% credible interval. This figure shows the final output gap estimate together with the 95% credible interval. Shaded areas highlight recessions as determined by the NBER Business Cycle Dating Committee.

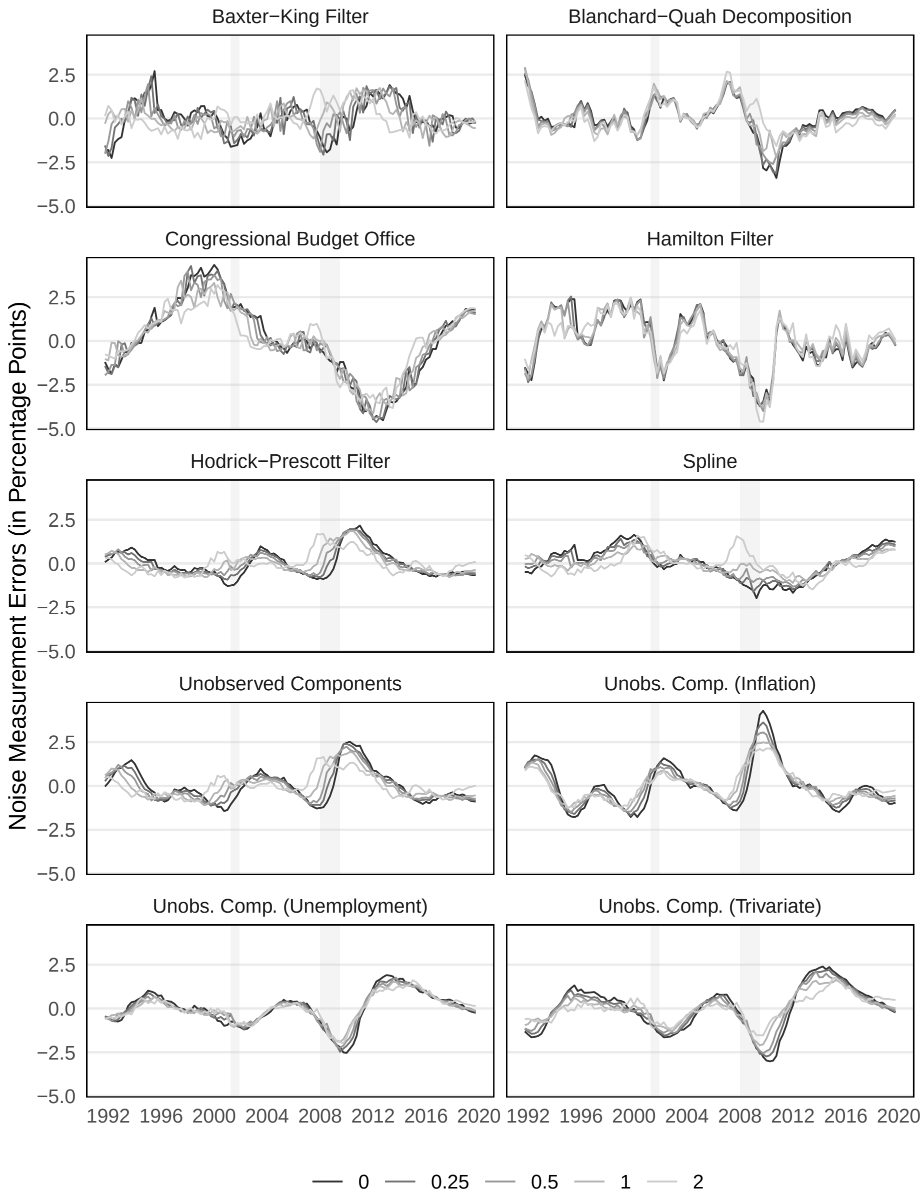


Figure 2.11: Noise measurement errors. This figure shows the estimated noise measurement errors for each output gap method using the full sample. The black lines depict the noise errors from the real-time release. The four gray lines do so for the subsequent releases. The lighter the grey, the later the release. Shaded areas highlight recessions as determined by the NBER Business Cycle Dating Committee.

Chapter 3

The Role of ECB Communication in Guiding Markets[1]

3.1 Introduction

In a remarkable 180-degree turn from their long-standing traditions of secrecy, many central banks today regard their external communication as an important tool for achieving their policy goals. Today's view is that communication can make monetary policy more effective by influencing expectations about monetary policy objectives and strategies, the economic outlook, and the (outlook for future) policy decisions (Blinder et al., 2008, 2017). Especially in a world in which key interest rates can hardly be lowered any further and new instruments to stimulate the economy are introduced, communication is seen as an important tool for influencing longer-term interest rates and yield curves. Hence, economists and central bankers nowadays agree that public perceptions about the stance of monetary policy is crucial for its effectiveness (De Haan and Sturm, 2019).[2]

[1]This chapter is joint work with Sina Streicher, Alexander Rathke and Jan-Egbert Sturm. It has been published in Public Choice (Anderes et al., 2021).

[2]In a survey conducted by Blinder et al. (2017), more than half of the responding central bank governors say that new communication measures have been adopted at their institution since the crisis. Moreover, the overwhelming majority (more than 80%) of central bankers and academic economists (more than 90%) stated that the role of central bank communication has intensified.

Since its existence, the European Central Bank (ECB) has used its main refinancing rate as the primary instrument of monetary policy. The Great Financial Crisis of 2008/2009 and the subsequent euro area debt crisis have caused the ECB to move this rate to such low levels that lowering it further became more and more difficult. Consequently, the ECB has started to deploy unconventional policy tools like asset purchase programs (APP) and targeted funding of bank lending to businesses and households. The tools were accompanied by forward guidance and intended to address the risks of a too prolonged period of low inflation. Accordingly, the ECB had to change its monetary policy communication strategy (Praet, 2013).[3]

The introduction of unconventional policy instruments has been highly controversial, especially among German economists, exemplified by presidents of the Bundesbank. They have been notorious for their opposition to any program involving the purchase of government debt.[4] According to the fundamental monetarist dictum, "inflation is always and everywhere a monetary phenomenon" caused by an increase in the money supply in excess of real output (Friedman, 1970). Ample evidence has identified a monetary overhang, especially when caused by financing governmental debts, as the cause of (hyper-) inflations (e.g., Bernholz, 2015). On the other hand, a collapsing money supply still is considered as the main cause of the Great Depression, which led to deflation (Friedman and Schwartz, 1963). Hence, the policies that lead to a huge increase in the money supply are still hotly debated. Bernholz (2015) states that "during the crisis such a dramatic increase of the monetary base ... took place as had formerly only been observed before high or hyperinflations. However, this time it happened without leading to inflation. On the other hand, broader money ... saw no unusual increase during the financial crisis. This leads to problems, namely which type of money could be responsible for later inflation".

Both the majority of central bankers (60%) and academic economists (75%) expect communication policy changes to remain in place, or to go even further (Blinder et al., 2017).

[3]As stated by Hartmann and Smets (2018), "the need for additional communication in a complex (non-standard) policy environment rose and forward guidance became an essential tool for easing policy in a low interest rate context".

[4]See, for instance, Jones (2014). For a critical view on unconventional policies, see Borio and Gambacorta (2017).

That controversy begs the question of how far markets perceived the actions of the ECB to be de- or inflationary.

Assessing the impact of those changes or summarizing the overall stance of monetary policy in such a different environment is challenging. To start with the latter, conventional monetary policy commonly is approximated by short-term policy interest rates (e.g., Mumtaz and Theophilopoulou, 2017; Coibion et al., 2017). In unconventional times, the primary instrument is constrained by the effective lower bound, making the assessment of the stance of monetary policy infeasible. A common way of circumventing that problem is the use of so-called shadow rates (Krippner, 2015; Wu and Xia, 2016; Inui et al., 2017). The shadow rate is derived from yield curve data and has no binding lower bound. In normal times, it approximately equals the policy rate, whereas in effective lower bound environments, it may turn negative reflecting movements at the far end of the yield curve. Considering longer maturity interest rates is essential for assessing the monetary policy stance; the ECB's unconventional tools like APP and forward guidance aim to influence exactly those rates. We follow Krippner (2013a,b, 2015) and measure the stance of monetary policy as interpreted by financial markets by using the Krippner shadow rate.

We want to gain insights into the extent to which financial markets react to the communication of the ECB. Further, we are interested in the relationship between communication and the expectations of professional forecasters.

Communication plays an extremely important role in the introduction and implementation of unconventional policy measures such as the quantitative easing (QE) programs. According to various studies, the strongest effects on asset prices can be felt precisely when such programs are announced. The expected consequences of the programs are therefore immediately reflected in asset prices. How the programs initially are "sold" could be a key factor in their success. Substantive evidence exists that ECB communication is able to reduce financial market variability (see, for instance, Coenen et al., 2017; Filardo and Hofmann, 2014). Moreover, financial markets are able to predict monetary policy decisions better when communication is collegial, i.e., by conveying the majority view of the committee instead of members' individualistic opinions (Ehrmann and Fratzscher, 2013). A broad consensus also exists that forward guidance moves financial markets in the intended direction (Brand et al., 2010; Ehrmann and Fratzscher, 2007; Galardo and Guerrieri, 2017; Swanson, 2017). That is also the case for the announcement of an APP, especially in the presence of forward guidance (Coenen et al., 2017). In line with the view that central banks manage expectations by using communication (Woodford, 2001), a second strand of literature has investigated the influence of communication on inflation expectations and, eventually, inflation outcomes. Empirical results suggest that transparency reduces the volatility of expectations and inflation (Van Der Cruijsen and Demertzis, 2007; Ullrich, 2008). Studies that quantify central bank communication differ in terms of the methodology applied. Some scholars measure communication by interpreting the tone of intermeeting speeches (Ehrmann and Fratzscher, 2007); others quantify the frequency of future-tense verbs in the first section of the introductory statement (Galardo and Guerrieri, 2017). Coenen et al. (2017) consider the length of the introductory statement, compute a language complexity index and classify different types of forward guidance and APP announcements to estimate their individual effects. Other strategies include the use of money market data to build communication indicators (Brand et al., 2010).

Our empirical analysis uses a unique dataset encompassing almost 43,000 classified statements that stem from the introductory statements after each ECB Governing Council meeting in which monetary policy decisions were made. We rely on human coding to accurately distil the information conveyed. In contrast to other studies that quantify communication, that method allows us to build more thorough and reliable communication measures that are consistent across time. Each press release is dissected into statements that subsequently are classified according to their topical association, time reference and qualitative content. The last reflects the tone ("increase", "stay the same", or "decrease"), allowing for an economic interpretation of the impact on various variables. From the statement-level data, we construct indicators that reflect key subjects in the introductory statement: price stability aspects, the economy and monetary developments. To measure the change in the market's interpretation of the stance of monetary policy, we compare the shadow rate the day after the Governing Council meeting to its value the day before. By concentrating on that short time window and furthermore controlling for, e.g., the previous shadow rate, the actual interest rate decision and its surprise component, the changes in expected and realized inflation, GDP and money (M3) growth, we intend to isolate the effects of the ECB's communicated actions and outlook on the market interpretation of monetary policy. Additionally, we want to assess the impact of ECB communication on the market's perspectives on exchange rates and inflation expectations. For that purpose, we consider the change in the nominal euro effective exchange rate and the change in inflation-indexed swap rates around ECB Governing Council meetings. In a second model, we analyze whether changes in professional forecasts, as measured by the consensus forecasts published by Consensus Economics, are related to the communicated assessments of the ECB Governing Council. We distinguish between expectations in inflation, GDP growth and M3 growth.

We find that ECB communication on the economy has been important in driving the shadow rate. This relationship appears stable over time, i.e., does not change across the different monetary policy phases we distinguish. When focusing on changes in the shadow rate, we therefore do not find evidence that communication has become more important. Communication on money is not immediately picked up by financial agents and thus not reflected in the shadow rate. Communication on money does appear to matter for inflation expectations as extracted from swap data. In our setup, exchange rates do not appear to be affected by ECB communication through its press releases. Its communication on price and in particular economic developments, however, does matter for the shadow rate. We find a significant, quantitatively important and stable impact of both on the assessment of financial market participants of the ECB's monetary policy stance, as measured by the shadow rate.

Not only financial market participants, but also professional forecasters appear to be influenced by the messages conveyed in the ECB's press releases. Messages related to the economy and price stability both impact changes in growth and inflation forecasts; those on money are related only to money growth forecasts. We thereby confirm the results already reported in Berger et al. (2011) that monetary developments in general appear to play only a minor role. The clear exception occurs during the period that covers both the Great Financial Crisis and the euro area crisis. During that period, actual increases in asset and security holdings related to monetary policy were associated with downward adjustments of growth forecasts; communication suggesting expansionary developments in the monetary sector coincided with downward forecasts in money growth rates. Hence, particularly during that crisis period these relationships appear to have been distorted.

The rest of the paper is structured as follows. We describe the different phases of monetary policy since the existence of the ECB in the next section. The data are introduced subsequently in Section 3.3. Section 3.4 presents the empirical results and Section 3.5 concludes.

3.2 Different phases of monetary policy

Since the start of its activities in 1999, the ECB has gone through various stages in its monetary policy operations. Clearly, policy was conducted in a different way before the onset of the Great Financial Crisis than it was afterwards. Those changes are documented in Figure 3.1, which depicts the evolution of the asset side of the ECB's balance sheet, the main refinancing rate and the shadow rate as a measure of the overall policy stance.[5]

Until the Great Financial Crisis, the shadow rate, in line with short-term money market rates, followed the path of the main refinancing rate closely. During that period, monetary policy was conducted almost exclusively by setting the three key ECB interest rates, which determined the path of money market rates. When the financial crisis hit, central banks around the world reacted in a consolidated fashion by providing an ample and stable supply of emergency liquidity to avoid a financial meltdown. They also cut policy rates decisively. Those efforts were most pronounced after the Lehman Brothers shock in September 2008. The ECB started to provide

[5]For a comprehensive treatment see, Hartmann and Smets (2018).

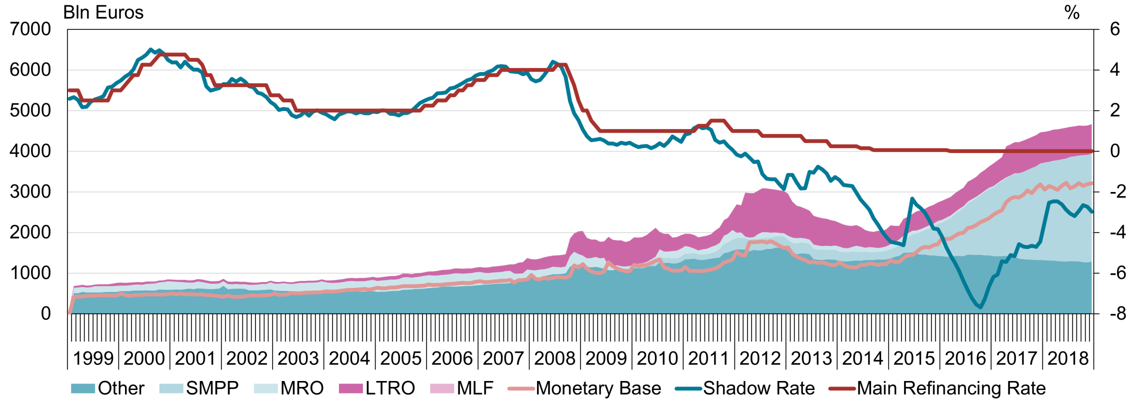


Figure 3.1: ECB's balance sheet. Notes: *SMPP* Securities held for monetary policy purposes, *MRO* main refinancing operations, *LTRO* longer term refinancing operations, *MLF* Marginal lending facility. Source: ECB, Krippner.

unlimited credit to banks at a fixed interest rate, an approach known as fixed-rate full allotment. Moreover, the maturity of refinancing operations was extended considerably and the range of eligible assets that could be used as collateral was expanded.

In 2010, the European Sovereign Debt crisis emerged, when the Greek fiscal situation deteriorated significantly and several other euro area countries subsequently became distressed. During the first stage of the crisis (until September 2012), the ECB bought a limited amount of troubled bonds under the Security Markets Program, but sterilized the effect on the monetary base using the Term Deposit facility. A policy of quantitative easing was not yet politically feasible. In contrast to other major central banks, the ECB started to raise rates in 2011, citing upside risks to price stability, which with the benefit of hindsight exacerbated the debt crisis unnecessarily. The shadow rate increased even before the official rate hikes, already indicating a more restrictive policy stance. When President Mario Draghi took over duties as ECB president in late 2011, policy reversed course. Monetary conditions were eased by two additional large liquidity-providing operations (VLTROs – very long-term refinancing operations[6]) and a later reversing of the interest rate hikes. Mario Draghi finally ended the debt crisis by giving his famous “Whatever it takes” speech on July 26, 2012, which was followed by the official announcement of the Outright Monetary Transactions (OMT) program that in principle allows the ECB to buy government bonds in unlimited quantity to avoid a self-reinforcing vicious feedback loop in sovereign bond markets. In fact, that particular program never was activated. Nevertheless, all of the foregoing measures led to a considerable easing of monetary conditions.

However, in 2013, the stance of monetary policy started to become more restrictive when banks repaid large fractions of money taken up under the VLTROs; the monetary base contracted significantly. The ECB began using explicit forward guidance in July 2013 when the ECB Governing Council said that it expected interest rates to remain low for an extended period of time. Significant worries about a continued downward trend in underlying inflation saw the ECB entering a new phase of monetary easing in June 2014. The Governing Council decided to introduce a negative rate on its deposit facility to overcome the zero lower bound problem, start new asset purchase programs for asset-backed securities (ABS) and covered bonds (CBPP) and to stop sterilizing asset purchases. In addition, the ECB introduced a facility to provide longer-term funding to banks for new loans, contingent on bank credit supply behavior, referred to as targeted longer-term refinancing operations (TLTRO).

Finally, in January 2015, as the last of the world’s major central banks, the ECB embarked on a large-scale asset purchase program (APP), which included sizable purchases of sovereign bonds. It was announced that the ECB would each month buy euro-area bonds from central governments, agencies and European institutions worth 60 billion euros. That program started in March 2015, and initially was supposed to last until September 2016. It was later extended in size and duration. Net purchases under the APP program finally ended in December 2018 and the

[6]Regular longer-term refinancing operations (LTRO) have maturities of three months. The VLTROS had a maturity of 36 months and were conducted as fixed rate full allotment procedures. They were accompanied by a reduction in the required reserve ratio from 2 to 1% and an increase collateral availability by broadening the definition of eligible assets.

ECB's balance sheet then contained assets held for monetary policy purposes worth 2.6 trillion euros.

Hence, we can summarize that, until the financial crisis monetary policy was conducted primarily through changes in policy rates, while in recent years the effect of unconventional measures dominated. Broadly in line with the ECB (Constâncio, 2018), we classify the monetary policy for the euro area into four phases. The first phase starts with the launch of the single currency and lasts until the revision of the monetary policy strategy in May 2003, when the weight of the monetary pillar and the "dominant role of money" were demoted, bringing the framework closer to the flexible inflation-targeting regime adopted by many other central banks around the world. Also, for purely technical reasons—we do not have expected inflation and growth rates for the euro area at large of a similar quality beforehand as we do afterward—we will ignore that period altogether in the rest of this paper.

The second phase spans from the revision of the monetary strategy to the outbreak of the Great Financial Crisis, i.e., the bankruptcy of Lehman Brothers in September 2008. The third phase then marks an abrupt change in euro area monetary policy. The fourth and final phase of monetary policy, the quantitative easing phase, starts in June 2014 with the announcement of the comprehensive package of expansionary measures described above.

3.3 Data

To summarize the overall stance of monetary policy, we use the shadow rate (SR) as developed by Krippner (2013a,b, 2015). Prior to the financial crisis in 2008, yield curve dynamics were often modelled using Gaussian affine term structure models (GATSM). One prominent downside of GATSMs is that they assign positive probabilities to negative interest rates. In other words, they do not account for the effective lower bound (ELB). Shadow models have become a popular tool to circumvent that problem. The shadow rate represents the short rate in a hypothetical cashless world. The difference between that hypothetical world and the real world is the opportunity to withdraw and store cash if interest rates turn negative. The feature is modelled by a put option, i.e., the right to exchange the shadow rate against the fixed ELB rate.[7] Hence, the short rate (R) is the sum of the shadow rate and the payoff from the put option: $R = SR + max\{\text{ELB - } SR, 0\}$. Consequently, whenever the shadow rate is (far) above the ELB, the option value becomes close to zero and the short rate coincides with the shadow rate. In contrast, in ELB environments, the SR may turn negative, reflecting movements at the long end of the yield curve.

In essence, Krippner's shadow model estimates the observed yield curve and the shadow yield curve simultaneously. To obtain the shadow rate, one essentially slides along the shadow yield curve to the point at which the term to maturity is zero.

Figure 3.2 depicts changes in the estimated euro yield curve and the corresponding shadow yield curve between two specific dates. The top panel illustrates the change between September 2008 and March 2009. During that period, the main refinancing rate (MRR) was reduced from 4.25 to 1.5% in five steps. While the shadow rate on September 19, 2008, is almost identical to the MRR, it is substantially lower on March 19, 2009, reflecting the rising value of the put option in near-ELB environments. At the beginning of 2018 (bottom panel), when shorter maturity rates hit the ELB, the increase at the long end of the yield curve from January 19, 2018, until February 19, 2018, affects the entire shadow yield curve, including the shadow rate. Thus, the SR offers a comprehensive indicator of monetary policy's stance across conventional and unconventional phases.

Krippner provides the shadow rate for the euro area on a daily basis, enabling us to assess the effect of policy decisions communicated by the ECB Governing Council on its meeting days.[8] Figure 3.1 shows the SR and the MRR as determined by the

[7]The idea dates back to Black (1995), who argued that when interest rates reach the zero lower bound, people would prefer to hold cash rather than financial instruments generating negative interest.

[8]The data produced by Leo Krippner can be accessed through `https://www.ljkmfa.com/visitors/`. In order to cope with the non-linearity introduced by the option effect, Krippner applies an Iterated Extended Kalman Filter for the estimation. He uses German and French zero government bond rates and overnight indexed swap rates (once they become available for the euro area in 2008) with maturities of 0.25, 0.5, 1, 2, 3, 5, 10 and 30 years to approximate the euro yield curve. For comparison, Krippner sets the ELB to 12.5 basis points across all estimated currencies. McCoy and Clemens (2017) analyze the effect of the choice of the effective lower

ECB Governing Council. Unsurprisingly, the SR and the MRR move closely together until the beginning of the European debt crisis. Both rates diverge sharply once the MRR enters the ELB environment. The contemporaneous correlation coefficient is 0.91.

We use real-time consensus forecasts as provided by Consensus Economics, Inc., to construct expected inflation, output growth and M3 growth series. The monthly survey includes estimates of prominent banks and forecast organizations on a range of variables such as future growth, inflation, interest rates and exchange rates. Each forecast is made for the current and the following year, enabling the construction of 12-month forecasts as a weighted average of both rates.[9] Since Consensus Economics did not publish forecasts for the euro area prior to January 2003, we restrict our estimation window to the 2003–2018 period.[10]

To reflect the latest economic information available, we collect real-time observations on actual inflation, GDP growth and M3 growth from the Real Time Database published by the ECB in its Statistical Data Warehouse.[11] For all three variables, we must realize that the "euro area" is a moving concept in the sense that the introduction of that currency in Greece, Slovenia, Cyprus, Malta, Slovakia, Estonia, Latvia and Lithuania redefines the area that is covered by our data, thereby implying benchmark revisions. Regarding inflation, we use the different vintages that are

bound in Krippner's shadow rate model estimation and give an illustrative example. They point out that, the higher the ELB is set, the more pronounced is the option effect, implying that the shadow yield curve (and, thus, the shadow rate) is shifted downwards. Hence, a trade-off exists between a meaningful monetary policy stance indicator and an economically relevant policy rate estimate. In our analysis, we use the shadow rate as a proxy for the stance of monetary policy and, therefore, are not concerned with the relatively high lower bound of 12.5 basis points.

[9] The weight for the forecast of the current year is $\frac{m}{12}$ and the weight for the forecast of the following year is $\frac{12-m}{12}$, where m denotes the number of remaining months in the current year. As the survey usually is published in the first half of the month, we assign the current month to the remaining months of the year.

[10] Using nominal GDP weighted averages from the euro area member states to construct expectations that go back to January 1999 does not change the conclusions. However, as money growth expectations are not available at the member state level, we would lose information by going back that far.

[11] The database can be accessed through `https://sdw.ecb.europa.eu/browse.do?node=9689716`. The first vintages of the for-us relevant variables published in that real-time database date from January 2001.

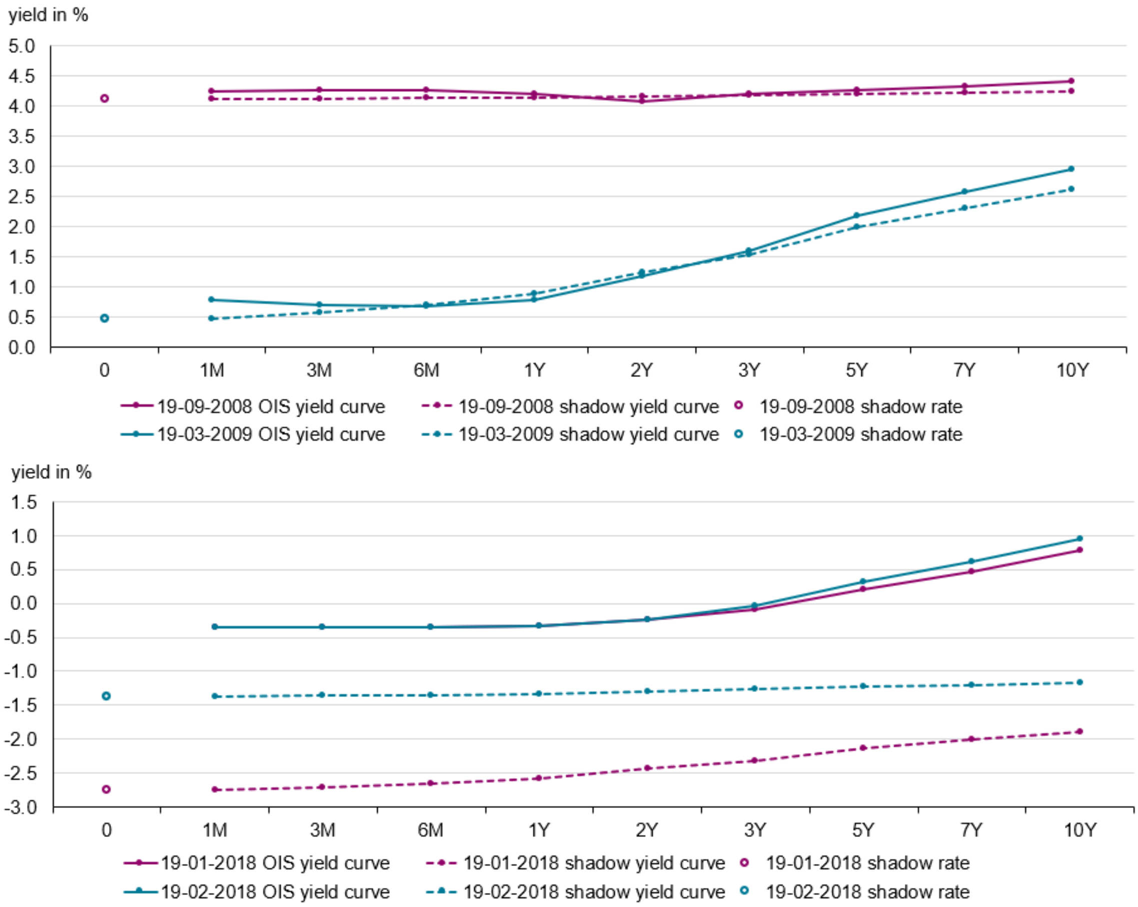


Figure 3.2: Euro Overnight Indexed Swaps (OIS) yield curve, estimated shadow yield curve, and shadow rate for specific dates prior to QE (top) and afterwards (bottom). Notes: The shadow yield curve is estimated from 3- and 6-month, and $1, 2, \ldots, 10$ year euro OIS rates. The euro OIS rates are taken from Datastream.

available for monthly year-over-year growth rates in the overall Harmonised Index of Consumer Prices (HICP). As to be expected, those inflation statistics hardly ever are revised. The small revisions that do occur take place within the first couple of releases or are required by other smaller benchmark revisions.[12] GDP growth is calculated from the chain-linked volumes of quarterly Gross Domestic Product. In

[12]The benchmark revisions took place early March 2002, June 2002, March 2003, May 2005, March 2011 and March 2016.

general, revisions to that variable tend to be larger and more frequent than those to the inflation variable. Revisions in monthly year-over-year growth in the monetary aggregate M3 are, on the other hand, more comparable to those in inflation.

The ECB's most important channel of communication is the ECB president's Introductory Statement following the Governing Council meeting (De Haan, 2008). Roughly once a month, the ECB Governing Council meets to discuss and take monetary policy decisions, which are announced publicly at 13:45 (CET). The subsequent press conference held by the ECB's president and vice president consists of two parts: a prepared Introductory Statement agreed upon on a word-by-word basis by all council members and a Questions & Answers session allowing journalists to address remaining questions. Beside the monetary policy decisions, the different sections of the Introductory Statement provide the ECB's assessment of developments in areas such as the real economy, prices and monetary aggregates.

In order to quantify ECB communication, we use indicators that capture the key subjects of the Introductory Statement for monetary policy, i.e., the development of the real economy, prices and monetary aggregates. The statements by the ECB president are analyzed by the media research institute Media Tenor. A media analyst codes each statement (several statements can be found within one sentence) by assigning a broad as well as a specific topic association, a time reference and its qualitative content. The latter is classified numerically into "increase", "stay the same", or "decrease". For instance, in the introductory statement from March 17, 2019, Mario Draghi said

Compared with the December 2016 Eurosystem staff macroeconomic projections, the outlook for real GDP growth has been revised upwards slightly in 2017 and 2018.

That sentence has been assigned to the broad topic "economy" and assessed to be an "increase". Overall, we use 42,827 statements contained in 168 Introductory Statements, each of which contains, on average, 255 coded statements. Our three ECB communication indicators (price stability, economy and money) are, on average, constructed from 37.9, 19.2 and 24.3 coded statements, respectively. The constructed indicators are the normalized balances of up- and downward assessments (including the neutral ones) regarding the specific topic. Formally,

$$\text{indicator}_{i,t} = \frac{\#\,\text{positive}_{i,t} - \#\,\text{negative}_{i,t}}{\#\,\text{positive}_{i,t} + \#\,\text{negative}_{i,t} + \#\,\text{neutral}_{i,t}}$$

where $i \in \{\text{price stability, economy, money}\}$. Accordingly, all indicators range theoretically from minus one to plus one.[13] Each of the panels in Figure 3.3 plots the evolutions of the statements' realized and expected counterparts, together with the respective communication indicator. Each data point reflects the information available at that particular moment in time. The resemblance between the communication indicators and their respective reference series clearly is visible.

In line with Bredin et al. (2010) and Fausch and Sigonius (2018), we use three-month Euribor future rates to construct the market's surprise reaction to the change in the main refinancing rate. Bernoth and Hagen (2004) show that Euribor future rates are unbiased and informationally efficient predictors of future spot rates. Consequently, we define the surprise component of the monetary policy decision by the change in the Euribor future rate from the end of the ECB Governing Council meeting day and the end of the previous day.[14] In order to control for the market's reaction to the US jobless claims published each Thursday, we use the difference between actual unemployment compensation claims and the median expected claims polled by Reuters on the previous Monday as a surprise indicator (Coenen et al.,

[13] The way in which our indicators are constructed very much follows the procedure underlying the so-called KOF Monetary Policy Communicator (MPC), as published by the KOF Swiss Economic Institute and used, e.g., by Sturm and De Haan (2011); Conrad and Lamla (2010); Lamla and Sturm (2013); Bulíř et al. (2013); Neuenkirch (2013). The key difference is that the MPC is a leading indicator for monetary policy and therefore considers only forward-looking statements regarding prices.

[14] The Euribor 3-month futures contracts have expiry dates on the 3rd Wednesdays of each of the following six months and at quarterly frequencies for longer horizons. Each contract can be traded until two days before the expiry date, meaning we do not have a natural series of future prices. We use an end-of-day series that represents the future contract that expires next and can still be traded. That series may contain price jumps once a month, exactly two days prior to the expiry dates. Owing to the timing of the ECB Governing Council meeting days, those jumps do not affect our market surprise measure.

2017). Hence, a positive surprise means that more claims were filed than expected previously.

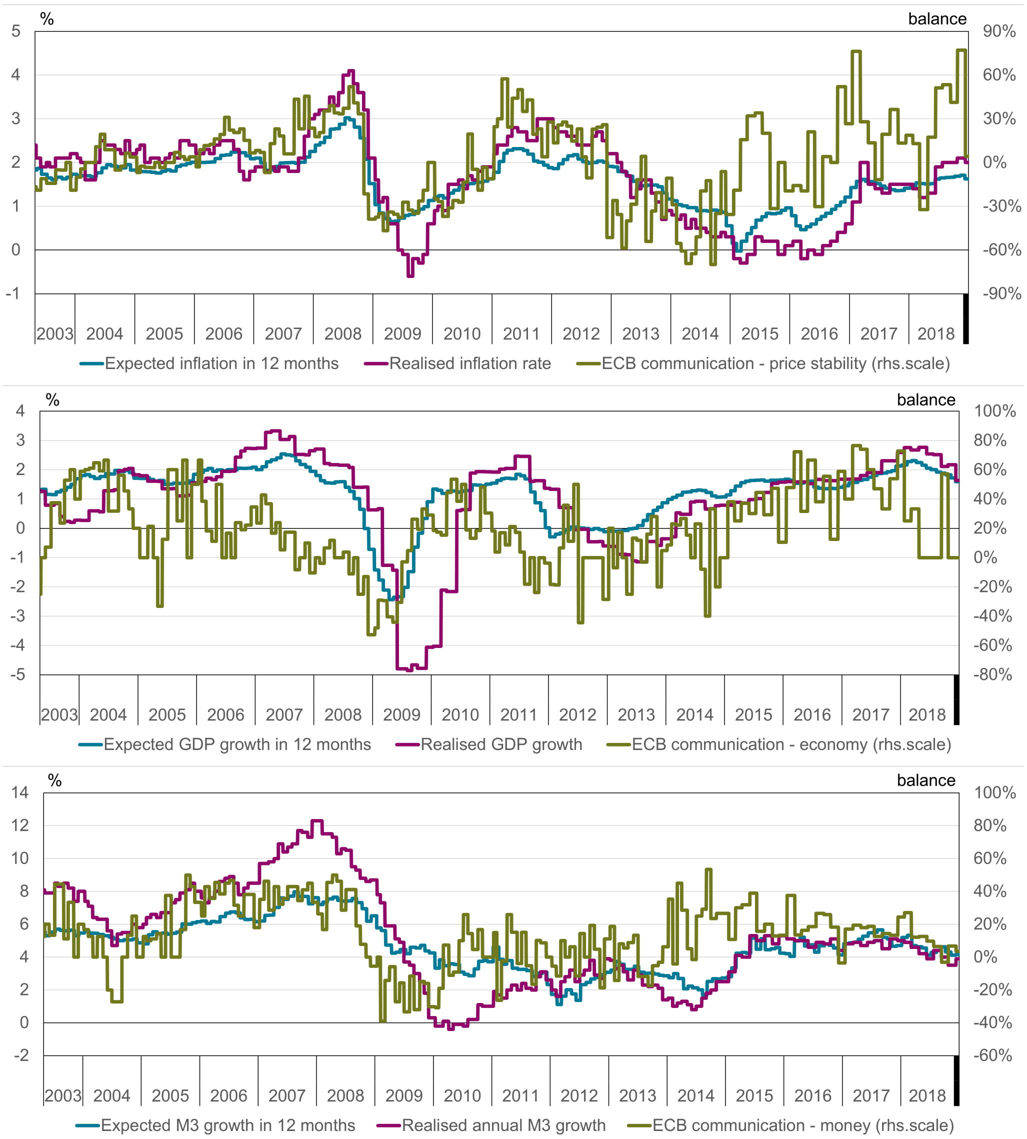


Figure 3.3: Forecast, realized and communicated information on inflation, economic growth and money growth, 2003m5 – 2018m12. Notes: The forecast information is taken from Consensus Economics, Inc. The realized data are taken from the real-time database of the ECB's Statistical Data Warehouse. The ECB communication data are extracted from ECB press releases.

As a robustness check, we also want to examine the effect of ECB communication on other measures that reflect financial market expectations. For that purpose, we examine both exchange rates and a financial market measure of inflation expectations. We use the euro's nominal effective exchange rate as published by the ECB to measure the foreign exchange market's perception of the value of euro relative to foreign currencies.Regarding inflation expectations, we use one-year euro inflation-linked swaps provided by Bloomberg on a daily basis, which have been available since 2004 (e.g., Grothe and Meyler, 2015). An inflation swap is a derivative contract that entitles the buyer to earn a fixed interest rate in exchange for a floating rate. The floating rate is linked to an inflation index. The fixed inflation swap rate therefore represents the market's inflationary expectation over a specific investment period. As in the case of the shadow rate, its daily frequency allows us to infer the immediate impacts of ECB communication on the change in market-based inflation expectations.

Table 3.1 reports summary statistics and contemporaneous correlation coefficients for all variables used. The largest contemporaneous correlations are found between the shadow rate and the main refinancing rate, between the expected and realized inflation rates and between expected and realized M3 growth rates.[15] The three different communication variables hardly are correlated with each other and, hence, appear to reflect substantially different informational contents of the ECB's press releases.

In the empirical analyses below, we either move to the frequency at which press releases after ECB Governing Council meetings are issued, or to the monthly frequency at which consensus forecasts are published. In the first case, we will concentrate on the change in the shadow rate around the Governing Council meetings, i.e., the difference in the shadow rate between the day before and the day after the ECB Governing Council's press conference. The analysis will be repeated for

[15]The change in the shadow rate and the change in the main refinancing rate are not correlated in this table because it takes up to 6 days for the main refinancing rate effectively to be changed after having been announced publicly by the ECB. By that time, the shadow rate already has adapted.

the change in the euro's nominal effective exchange rate and the change in inflation expectations based on swaps.

Given the different monetary policy phases summarized in Section 3.2, we are interested not only in analyzing the impact of ECB communication during the full sample period, but also whether we can observe structural changes across those phases. In a first step, Table 3.2 reports the correlation coefficients between the change in the shadow rate, on the one hand, and the actual decided change in the main refinancing rate as well as the three communication indicators based on the ECB press releases, on the other hand.[16] What catches the eye is that the largest correlation is between the ECB communication indicator reflecting price stability and the change in the shadow rate. However, that holds only for the last monetary policy phase we distinguish. Whereas Figure 3.1 suggests a close link between the main refinancing rate and the shadow rate until the financial crisis. The correlation between the change in each of them is quite stable across the different phases. Overall, the ECB's communicated assessment of the economic situation and outlook is most strongly correlated with changes in the shadow rate.

Another way in which we can split our sample is by comparing each interest rate decision with what would have been implied by a rule-based approach. Following Cochrane et al. (2019), we construct a measure indicating whether the actual decision aligns with a (Taylor) rule-based policy and check whether the effect of communication on expectations depends on it. When the interest rate decision is in line with a Taylor rule, the correlation between the change in the MRR and the change in the SR is almost non-existent. The same conclusion, albeit to a much lesser extent, also applies to our surprise measure and the shadow rate. Apparently, market participants are more likely to have anticipated policy decisions that are consistent with a Taylor rule. Communication regarding the economy appears to be especially relevant when the ECB deviates from what a Taylor rule would have implied.

[16]The effective change in the MRR usually takes place six days after it was announced publicly. Since financial markets incorporate the announced change in the MRR immediately, we use the decided change in MRR in the subsequent analyses.

						Correlation coefficients													
		Mean	SD	Min	Max	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1)	Shadow Rate (%)	-0.50	3.11	-7.69	4.44	1													
(2)	Nominal Effective Exchange Rate (index)	101.59	5.42	88.16	114.44	0.76	1												
(3)	Inflation expectations based on swaps (%)	1.45	0.68	-0.76	3.49	0.71	0.47	1											
(4)	Main refinancing rate (MRR) (%)	1.36	1.31	0.00	4.25	0.91	0.70	0.65	1										
(5)	3-Month Euribor future (%)	1.29	1.61	-0.34	5.26	0.89	0.66	0.66	0.99	1									
(6)	Surprise in US Jobless Claims (%)	0.10	4.53	-14.29	23.65	0.12	0.09	0.08	0.12	0.12	1								
(7)	Expected inflation in 12 months (%)	1.57	0.60	-0.02	3.02	0.68	0.42	0.81	0.69	0.68	0.10	1							
(8)	Realised inflation rate (%)	1.61	1.03	-0.60	4.10	0.64	0.38	0.75	0.62	0.62	0.09	0.94	1						
(9)	Expected GDP growth in 12 months (%)	1.21	1.00	-2.44	2.54	0.05	-0.21	0.28	0.15	0.21	0.02	0.25	0.16	1					
(10)	Realised GDP growth (%)	1.03	1.73	-4.85	3.33	0.08	-0.24	0.28	0.23	0.31	-0.01	0.40	0.41	0.72	1				
(11)	Expected M3 growth in 12 months (%)	4.60	1.59	1.11	7.98	0.49	0.37	0.40	0.70	0.73	0.06	0.40	0.31	0.40	0.41	1			
(12)	Realised annual M3 growth (%)	4.92	3.11	-0.40	12.30	0.50	0.29	0.41	0.73	0.77	0.04	0.46	0.39	0.29	0.49	0.90	1		
(13)	ECB communication - price stability (balance)	0.03	0.30	-0.70	0.77	0.08	-0.11	0.45	0.18	0.24	0.05	0.48	0.46	0.43	0.52	0.33	0.32	1	
(14)	ECB communication - economy (balance)	0.18	0.27	-0.53	0.76	-0.33	-0.39	-0.05	-0.26	-0.23	-0.05	-0.17	-0.24	0.48	0.18	0.08	-0.06	0.11	1
(15)	ECB communication - money (balance)	0.13	0.20	-0.39	0.53	0.12	-0.15	0.19	0.29	0.36	0.02	0.20	0.14	0.56	0.58	0.45	0.53	0.24	0.15

Table 3.1. Summary statistics and contemporaneous correlation coefficients. Notes: The data cover the period May 1st, 2003, until December 31st, 2018.

	Full period	Phase 2 2003:5 −2008:8	Phase 3 2008:9 −2014:5	Phase 4 2014:6 −2018:12	Not rule -based	Rule -based
Decided change in main refinancing rate	0.20	0.15	0.21	0.17	0.31	0.08
Surprise in MRR (based on euribor futures)	0.56	0.77	0.65	0.13	0.65	0.49
Surprise in US Jobless Claims	−0.15	−0.17	−0.21	−0.08	−0.13	−0.16
ECB communication – price stability	0.11	−0.07	−0.05	0.38	0.11	0.13
ECB communication – economy	0.28	0.27	0.28	0.22	0.45	0.12
ECB communication – money	0.11	−0.01	0.12	0.01	0.05	0.18
Observations	168	60	69	39	73	95

Table 3.2. Correlation of the change in the shadow rate with the change in the ECB main refinancing rate and all three communication indicators. Notes: The frequency has been reduced to the frequency at which press releases after Governing Council meetings took place. The May 2003 – December 2018 period is covered, leaving 168 observations. Each cell shows the correlation coefficient with the change in the shadow rate.

3.4 Results

3.4.1 Explaining the immediate impact on the shadow rate

In our first model, we want to explain the reactions of financial market participants as measured by changes in the shadow rate on both the decisions made by the ECB Governing Council and its communication. We use the change in the shadow rate between the day before and the day after the ECB Governing Council meetings that are followed by a press conference. By concentrating around those meeting days, we reduce the probability that developments other than the ECB's explicit decisions and communication are driving the movements in the shadow rate.

The most general version of our model is summarized by

$$\tilde{\Delta} s_{t_i} = \alpha s_{t_i - 1} + \beta ECB\ communication_{t_i} + \gamma z_{t_i}^{market} + \delta z_{t_i}^{economy} + \varepsilon_{t_i} \qquad (3.1)$$

where t_i is the day of the i-th ECB Governing Council meeting in our sample, the dependent variable $\tilde{\Delta} s_{t_i} = s_{t_i+1} - s_{t_i-1}$ denotes the change in the shadow rate the day before and the day after the i-th meeting, and $z_{t_i}^{market}$ includes financial market control variables such as the decided and surprise change in the main refinancing rate as well as the surprise in US jobless claims. The economic control variable $z_{t_i}^{economy}$ includes realized and expected twelve-month inflation rates, GDP growth and M3 growth. The variable $ECB\ communication_{t_i}$ includes our three ECB communication measures on price stability, the economy and money, respectively. The exact timing of the model is illustrated in Figure 3.4 in the appendix.

The first column in Table 3.3 indeed reveals a positive and significant correlation between the decided change in the main refinancing rate and the movement of the shadow rate. Despite its significance, the size of the coefficient is at first glance small. Only 7% of the change in the main refinancing rate is reflected immediately in the shadow rate. That result is, however, not that surprising, given that most of the time markets are well prepared regarding upcoming interest rate decisions and, hence, most likely already have incorporated them in the shadow rate the day before the Governing Council meeting. The same conclusion also is supported by the next column showing that the shadow rate reacts much more strongly to the surprise component of the interest rate decision. When we use our three different communication measures to explain movements in the shadow rate, the results in column (3) show that, although all three have the expected positive signs, only that part of the press release that indicates the views of the ECB regarding the direction in which the real economy is evolving has a significant impact on the change of the shadow rate.[17] Comparing columns (1) and (3) highlights that the information revealed in the press release is more salient to the financial markets than the mere interest rate decision itself: albeit still small, the adjusted R2 more than doubles in size. Including the initial value of the shadow rate does not change things (column (4)). When we enter the decision and the surprise variable next to our communication measures, we still find an improvement in the adjusted R2 and, again, in our communication variables; only the one on the economy has a significant impact on the change in the shadow rate (column (5)).

[17] Including each of these communication variables separately does not change the conclusion: only communication on the economy is significant in explaining the change in the shadow rate. Given the low correlation between those indicators, that was to be expected (see Table 3.2).

In column (6), we add several other variables that might be related to the monetary policy decision and thereby perhaps to the change in the shadow rate. They include changes in professional forecasts, but also in releases of "hard" data regarding inflation, GDP growth and M3 growth since the previous Governing Council meeting. The only variable that helps to explain changes in the shadow rate further is the change in the expected GDP growth rate. When expected GDP growth picks up, the shadow rate tends to pick up as well. Entering expected GDP growth, however, does not change the significance of the coefficient capturing the impact of ECB communication on the economy. As to be expected, the size of the coefficient, however, is modestly reduced further. Once we incorporate the additional control variables, the actual interest rate decision no longer has a significant impact on the shadow rate.[18]

In a next step and based upon the model's full specification in column (6), we apply a general-to-specific methodology while keeping all communication variables on board.[19] The results are shown in column (7). The change in the consensus forecast of GDP growth and the interest rate surprise remain significant. In addition to the ECB's communication on the economic situation and the outlook, communication on inflation now also has a positive and significant impact on the change in the shadow rate. That does not change if we no longer restrict the general-to-specific approach by including all communication variables (column (8)).[20] How important is the impact of ECB communication on the real economy?[21] A one standard deviation change in overall ECB communication leads roughly to a one-quarter standard deviation movement in the shadow rate. An effect of comparable magnitude is found for changes in expected GDP growth. A similar variation in the interest rate surprise

[18] All results in Table 3.3 are based upon those days in which a Governing Council meeting has taken place and a monetary policy press release has been published. If we move to a daily frequency, increasing the size of the sample to 5724 observations, the significance of the variables increases, while no significant changes in signs are recorded.

[19] A general-to-specific methodology means dropping insignificant variables one at a time until all variables remaining in the model are significant.

[20] We have furthermore carried out a placebo test in which the dependent variable has been moved to another arbitrary date and find that in general all variables become insignificant.

[21] Standardized beta coefficients are shown in Table 3.9 in the appendix.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Shadow Rate at day before meeting				-0.000586	-0.000534	-0.000150		
				(-0.436)	(-0.431)	(-0.133)		
ECB communication – price stability			0.0117	0.0123	0.0126	0.0126	0.0195**	0.0221**
			(0.919)	(0.957)	(1.201)	(1.225)	(2.149)	(2.353)
ECB communication – economy			0.0439***	0.0423***	0.0301***	0.0251**	0.0232**	0.0252**
			(3.077)	(3.118)	(3.035)	(2.373)	(2.146)	(2.362)
ECB communication – money			0.0125	0.0139	0.00715	0.00390	0.0169	
			(0.817)	(0.846)	(0.476)	(0.285)	(1.238)	
Decided change in main refinancing rate	0.0634***				0.0500***	0.0256		
(change since day before meeting)	(3.472)				(3.179)	(1.526)		
Surprise in MRR (based on euribor futures)		0.721***			0.725***	0.704***	0.745***	0.745***
(change since day before meeting)		(7.152)			(7.932)	(8.027)	(8.534)	(8.419)
Surprise in US Jobless Claims						-0.000770		
						(-1.604)		
Expected inflation in 12 months						0.00360		
(change since previous meeting)						(0.187)		
Expected GDP growth in 12 months						0.0350**	0.0464***	0.0445***
(change since previous meeting)						(2.420)	(4.281)	(4.060)
Expected M3 growth in 12 months						-0.0128		
(change since previous meeting)						(-1.499)		
Realised inflation rate						0.0134		
(change since previous meeting)						(1.173)		
Realised annual GDP growth						0.000145		
(change since previous meeting)						(0.0188)		
Realised annual M3 growth						0.0121		
(change since previous meeting)						(1.553)		
Constant	-0.00231	-0.00327	-0.0129**	-0.0128**	-0.00901**	-0.00780**	-0.00997***	-0.00813**
	(-0.655)	(-1.103)	(-2.426)	(-2.415)	(-2.435)	(-2.003)	(-2.744)	(-2.548)
Observations	168	168	168	168	168	168	168	168
Adjusted R-squared	0.034	0.312	0.070	0.066	0.392	0.425	0.415	0.414

Table 3.3. Explaining the change in the shadow rate around Governing Council meetings. Notes: The sample covers the May 2003 – December 2018 period. Robust t-statistics in parentheses.
***p < 0.01, **p < 0.05, *p < 0.1.

leads to more than half of a standard deviation change in the shadow rate. Hence, only the interest surprise variable appears quantitatively more important than, in particular, the messages regarding price stability and the economy as contained in the ECB's press releases.

How stable are the detected relationships across the different monetary policy phases identified in Section 3.2? To answer that question, we allow all variables identified in column (7) of Table 3.3 to differ in phases 3 and 4.[22] For that purpose, we introduce dummy variables reflecting each of those phases and interact them with all included variables. The regression results are displayed in Table 3.4. As the individual dummy and interaction coefficients for the communication variable

[22]Hence, we use phase 2 as our baseline. As noted, phase 1, owing to data limitations, is not included in our sample. Similar analyses have been carried out for the other columns in Table 3.3. The conclusions are not affected.

suggest, and as is formally shown by p-values of joint F-tests on them, no statistical differences across the phases are found regarding any of the variables. In addition, when computing an overall Chow test, of which the p-value is shown in the last row in the left panel of Table 3.4, that conclusion does not change. Therefore, the shadow rate response seems remarkably stable over such different periods. Complementary to our policy phases, we classify each ECB Governing Council meeting day into rule-based or not rule-based phases and then estimate a similar regression model with dummy interaction terms according to those classifications. We follow Cochrane et al. (2019) and Nikolsko-Rzhevskyy et al. (2014) and apply an adjusted Taylor rule that accounts for the zero lower bound on interest rates in order to allocate each policy decision to the two phases. Whenever the difference between the policy rate and the Taylor-rule-implied rate declines, a meeting is classified as rule-based; if the difference increases, it is classified as not rule-based. The results are presented in the right panel of Table 3.4. Looking at the individual coefficients across the two phases, we again do not find evidence of structural breaks. In addition, an overall Chow test (as reported in the last row of the right panel in Table 3.4) cannot reject the null hypothesis of no significant differences across them. By focusing on the immediate impacts of the press release and monetary policy decision on the shadow rate, we have so far found significant, strong, stable and robust relationships between those aspects of the press release that interpret the real economy and price stability through the ECB's eyes.

3.4.2 Do other financial market variables react to ECB Communication?

In an additional exercise, we swap the change in the shadow rate for the change in the nominal effective exchange rate and the change in market-based inflation expectations, respectively. The results are presented in Table 3.5. In a first regression, we include all ECB communication indicators and all control variables; we then use the same general-to-specific procedure as before. Regarding the change in the nominal effective exchange rate, we observe that none of the communication indicators has a significant impact. The only variables that affect the change in the nominal effective

	Phase 2	ΔPhase 3	ΔPhase 4	Not rule-based	ΔRule-based
ECB communication – price stability	0.00552	-0.00544	0.0571	0.0334***	-0.0201
	(0.241)	(-0.205)	(1.551)	(2.611)	(-1.096)
ECB communication – economy	0.0192	0.00272	0.00645	0.0363**	-0.0276
	(1.476)	(0.0992)	(0.267)	(2.220)	(-1.299)
ECB communication – money	0.0186	-0.00560	0.0545	-0.00439	0.0364
	(1.253)	(-0.197)	(0.414)	(-0.206)	(1.253)
Surprise in MRR (based on euribor futures)	0.721***	0.0514	-0.229	0.858***	-0.216
(change since day before meeting)	(9.666)	(0.320)	(-0.979)	(7.404)	(-1.344)
Expected GDP growth in 12 months	0.0457	-0.00356	0.0367	0.0491***	-0.0104
(change since previous meeting)	(1.636)	(-0.112)	(0.470)	(3.730)	(-0.484)
Constant	-0.00863	-0.00305	-0.0154	-0.0111*	0.00495
	(-1.529)	(-0.411)	(-0.621)	(-1.678)	(0.637)
Observations		168		168	
Adjusted R-squared		0.430		0.417	
F-test phases com. prices equal (p-value)		0.097			
F-test phases com. economy equal (p-value)		0.449			
F-test phases com. money equal (p-value)		0.874			
F-test phases communication equal (p-value)		0.038			
F-test phases MRR equal (p-value)		0.063			
F-test phases exp. GDP gr. equal (p-value)		0.317			
Chow test (p-value)		0.00000834		0.209	

Table 3.4. Differences across phases or decision type. Notes: The sample covers the May 2003 – December 2018 period. Robust t-statistics in parentheses.
***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

exchange rate significantly around ECB Governing Council meeting days turn out to be the decided change as well as the surprise in the MRR, along with expected M3 growth.

However, ECB communication regarding money does have a positive and significant effect on the change in inflation expectations. In a quantitative sense, it has the strongest impact of all of the variables we consider. According to Table 3.9 in the appendix, which reports the standardized results of the general-to-specific approach, a one standard deviation change in the communication variable regarding money leads to approximately more than a fifth of a standard deviation change in inflation expectations. The surprise in US jobless claims as well as in expected M3 growth have negative and significant effects, albeit they are quantitatively less strong than that of our money-communication variable. Unsurprisingly, the expected twelve-month ahead inflation rate has a significant and positive impact on inflation expectations, although it also is quantitatively less strong than that of ECB communication regarding money.

We can conclude that while we do not find evidence that ECB communication affects exchange rates, ECB communication on monetary phenomena does affect inflation expectations based on swaps. At least in that market, participants consider ECB communication related to its monetary pillar to be indicative for future inflation. That exchange rates are unrelated to ECB communication is consistent with the claim of ECB officials that exchange rates are not an ECB policy target and are in general not commented on.

	Nominal effective exchange rate			Inflation expectation (swaps)		
	Full	Gen-to-Spec	Gen-to-Spec	Full	Gen-to-Spec	Gen-to-Spec
Shadow Rate at day before meeting	0.000236			-0.0108		
	(0.0228)			(-1.080)		
ECB communication – price stability	-0.187	-0.00470		0.0103	0.00569	
	(-1.054)	(-0.0324)		(0.472)	(0.282)	
ECB communication – economy	0.116	0.119		0.00881	0.0313	
	(0.703)	(0.863)		(0.363)	(1.269)	
ECB communication – money	0.225	0.231		0.0921***	0.0832***	0.0844***
	(1.014)	(1.264)		(2.782)	(3.049)	(3.172)
Decided change in main refinancing rate	-0.705**	-0.598***	-0.479**	0.0426		
(change since day before meeting)	(-2.458)	(-2.723)	(-2.364)	(0.924)		
Surprise in MRR (based on euribor futures)	6.804***	7.056***	7.150***	0.437*		0.397*
(change since day before meeting)	(5.850)	(6.151)	(6.392)	(1.778)		(1.721)
Surprise in US Jobless Claims	-0.00769			-0.00265*	-0.00355**	-0.00289*
	(-0.708)			(-1.756)	(-2.042)	(-1.857)
Expected inflation in 12 months	0.773			0.0835		0.0875*
(change since previous meeting)	(1.191)			(1.594)		(1.833)
Expected GDP growth in 12 months	-0.280			0.0138		
(change since previous meeting)	(-0.921)			(0.465)		
Expected M3 growth in 12 months	0.314**	0.298**	0.337**	-0.0346**	-0.0325**	-0.0309**
(change since previous meeting)	(2.352)	(2.206)	(2.590)	(-2.480)	(-2.275)	(-2.227)
Realised inflation rate	0.196			-0.0248		
(change since previous meeting)	(1.293)			(-0.856)		
Realised annual GDP growth	-0.0471			0.00257		
(change since previous meeting)	(-0.644)			(0.230)		
Realised annual M3 growth	-0.0182			-0.00516		
(change since previous meeting)	(-0.180)			(-0.387)		
Constant	-0.113	-0.0926*	-0.0389	0.00824	-0.0110	-0.00572
	(-0.104)	(-1.727)	(-0.974)	(0.501)	(-1.326)	(-0.914)
Observations	168	168	168	155	155	155
Adjusted R-squared	0.231	0.227	0.232	0.099	0.082	0.128

Table 3.5. Explaining changes in the exchange rate and market-based inflation expectations. Notes: The sample covers the May 2003 – December 2018 period. Robust t-statistics in parentheses.
***p < 0.01, **p < 0.05, *p < 0.1.

3.4.3 The impact on the expectations of professional forecasters

The above subsection provided evidence that financial market participants listen and react to what the ECB communicates. The shadow rate and inflation expectations react within a day to messages conveyed in the ECB's press releases. In a next step, we investigate the extent to which professional forecasters are affected by both monetary policy decisions and the information contained in the associated press releases of the ECB. For that purpose, we look at changes in the twelve-month ahead consensus forecasts constructed from monthly forecast surveys published by Consensus Economics, Inc. In line with our communication indices, we concentrate on inflation, GDP growth and M3 growth projections for the euro area. As before, the sample starts in May 2003 and ends in December 2018.[23]

A general version of our second model is given by[24]

$$\Delta\hat{x}_{k,t_j} = \alpha\hat{x}_{t_{j-1}} + \beta ECB\ communication_{t_j} + \gamma z_{t_j}^{ECB} + \delta z_{t_j}^{economy} + \varepsilon_{t_i} \qquad (3.2)$$

where t_j is the day of the j-th consensus forecast release date in our sample and the dependent variable $\Delta\hat{x}_{k,t_j} = \hat{x}_{k,t_j} - \hat{x}_{k,t_{j-1}}$ denotes the change of the twelve-month ahead consensus forecast of variable k between t_j and its previous release t_{j-1}.

We distinguish four sets of explanatory variables. The first set contains the consensus forecasts at the beginning of the period ($\hat{x}_{t_{j-1}}$). We expect to see some mean reversion in our data, i.e., strong (weak) forecasts are likely subsequently to be reduced (increased).

The second set represent our communication variables. $ECB\ communication_{t_j}$ denotes the latest available version of our three communication variables prior to the j-th release of the consensus forecast. Our expectation is that statements suggesting improvements or deteriorations in the to-be-forecasted economic variable will

[23]This time, however, we are working with what boils down to a monthly frequency that did not change over time.
[24]An illustration of the timing of the model can be found in Figure 3.5 in the appendix.

lead forecasters to adjust their expectations for that particular variable in the same direction.

The third set includes variables associated with ECB actions ($z_{t_j}^{ECB}$). As in the previous section, we enter the change in the main refinancing rate and the interest rate surprise derived from money market futures at decision days. In addition, we include the realized change in total assets and securities on the ECB's balance sheet that reflect monetary policy actions.[25][26] The last variable group measures changes in our variables of interest as reported officially and available at the time of the consensus forecast ($z_{t_j}^{economy}$). It includes the latest version of realized inflation rates, GDP growth and M3 growth published prior to t_j. We expect that official information released regarding the forecasted variables will influence the expectations of the professional forecasters.

Table 3.6 reports summary statistics for all variables just described. It reveals that the average inflation, economic growth and money growth forecasts in our sample hover around 1.6%, 1.2% and 4.7%, respectively. The changes in those forecasts are, on average, close to zero.

Table 3.7 reports the results for three alternative specifications. The first column for each outcome variable includes all variables described and available. The second column contains the results from a simplified model where variable reduction is based on the same general-to-specific methodology as before. Finally, the three third columns are estimated jointly using the seemingly unrelated regression method. That method allows for a potential gain in efficiency, as the residuals across these three equations are likely to be correlated.[27]

The results of the first set of variables reveal that, in general, forecasts revert to the mean. Such reversion is most pronounced for GDP growth forecasts. For money growth, however, we hardly can reject the hypothesis that we are observing

[25]Note that it did not make much sense to include this variable in our first model when we looked at the changes between the day before and the day after the Governing Council meeting. In such a short time frame, including it always resulted in insignificant coefficient estimates.

[26]Since the Great Financial Crisis, the variable takes on non-zero values. Hence, it covers only the last two monetary policy phases of our sample, i.e., 2008m9–2014m5 and 2014m6–2018m12.

[27]The correlation between the residuals of the inflation and GDP growth equations is the highest and equals 0.27.

	Obs.	Average	SD	Min	Max
Change in 12-months ahead inflation forecast (in %-points)	188	−0.0011	0.1126	−0.5539	0.2480
Change in 12-months ahead GDP growth forecast (in %-points)	188	0.0018	0.1810	−0.7173	0.8379
Change in 12-months ahead M3 growth forecast (in %-points)	188	−0.0074	0.3421	−0.9344	0.9530
Previous 12-months ahead inflation forecast (in %)	188	1.5834	0.5773	−0.0233	3.0235
Previous 12-months ahead GDP growth forecast (in %)	188	1.2283	0.9724	−2.4352	2.5372
Previous 12-months ahead M3 growth forecast (in %)	188	4.6699	1.5562	1.1092	7.9767
ECB Communication – price stability	188	0.0204	0.2966	−0.7000	0.7714
ECB Communication – economy	188	0.2002	0.2755	−0.5263	0.7647
ECB Communication – money	188	0.1344	0.1972	−0.3913	0.5333
Change in the Main Refinancing Rate (in %-points)	188	−0.0133	0.1421	−1.0000	0.2500
Surprise in MRR based on 3m Euribor futures (in %-points)	188	0.0010	0.0359	−0.1250	0.1550
Change in Assets & Securities for monetary policy purposes (in bln €)	188	14.0998	25.5838	−8.0810	102.6400
Change in the realised inflation rate (in %-points)	188	−0.0016	0.2593	−1.1000	0.9000
Change in the realised GDP growth rate (in %-points)	188	0.0047	0.4470	−3.3701	2.7583
Change in the realised M3 growth rate (in %-points)	188	−0.0245	0.4905	−1.5000	1.3000

Table 3.6. Summary statistics of variables used in explaining changes in forecasts. Notes: All variables cover the period May 2003 – December 2018.

a random walk process. Furthermore, money growth forecasts appear unrelated to previous forecasts in inflation and GDP growth. According to the general-to-specific results, the same holds for inflation expectations. Previous GDP or money growth expectations are not associated with subsequent changes in inflation expectations. Changes in growth expectations are, on the other hand, also affected negatively by high inflation expectations at the previous consensus survey release.

The impact of the communication variables mostly is significant and qualitatively at least as important as that of the first set of variables. Whereas ECB communication related to money is helpful only in explaining changes in money growth forecasts, communication on price stability and the economy is significant in explaining both changes in inflation and growth expectations. As expected, communication regarding the course of the economy is more important in explaining growth than inflation forecasts. The same holds true for the importance of price-stability communication on inflation versus growth forecasts. As the standard deviations of both communication variables are of about the same magnitude, the relative sizes of the coefficient estimates can be compared directly.[28] Accordingly, when explaining inflation forecasts, the impact of communication on price stability is almost twice as large as that of communication on the economy, and vice versa.

The changes in the main refinancing rate and in the assets and securities linked directly to monetary policy are both robust in explaining forecast revisions of GDP growth and inflation. Neither of those two measures of actual policy implementation helps in explaining changes in money growth forecasts. Looking at the estimated signs of those variables reveals that a loosening of monetary policy is associated with both lower growth and inflation forecasts for the upcoming 12 months. Instead of boosting these outlooks, we find that professional forecasters rather interpret expansive policies as signs of deteriorating economic conditions. Note that the foregoing effects are conditional on the communicated assessment of the ECB and recently published hard data. Hence, even if we hold official data and ECB communication constant, actions that loosen monetary policy are associated with downward revisions of growth and inflation forecasts using a forecast horizon of 12 months.

Finally, it comes as no surprise that the release of official data drives forecast revisions in the same direction. The estimation results suggest, however, that only the data directly related to the variable forecasted has an impact on those forecasts, i.e., the realized inflation rate helps explain inflation forecasts, the GDP growth rate the GDP growth forecast and M3 growth the growth forecast for M3.

Taken together, the communication variable measuring price stability has the quantitatively largest impact on inflation forecasts. A one standard deviation change in that communication variable translates into more than a third of a standard deviation change in the inflation forecast. Although the impact of communication regarding the economy on the GDP growth forecast is even larger, at 0.4 standard deviations, both the initial forecasts and the change in the main refinancing rate are at least as important quantitatively. Our model can explain approximately half of the variation in changes in growth forecasts. Changes in money growth

[28]Table 3.10 in the appendix reports standardized coefficients of the results presented in Table 3.7, allowing for a direct comparison of the quantitative importances of each of the explanatory variables.

forecasts are much more difficult to model. The only three significant variables each have a quantitative impact of about 20% of a standard deviation triggered by a one standard deviation change (see Table 3.10 in the appendix). The resulting R2 merely is around 0.1.

How robust are those results across the three different monetary policy phases that we distinguish? In Table 3.8, we first report results that allow all ECB variables to differ across the phases. Subsequently, we apply the same general-to-specific methodology as before. At the bottom of the table, we report *p*-values of F-statistics testing the null hypothesis that the respective interaction terms are zero, implying no statistically significant differences across the monetary policy phases.

In contrast to our first model in which we explained changes in the shadow rate and did not find any notable differences across these monetary policy phases, we now find that the ECB communication variables do have statistically different impacts across the phases when it comes to adjustments in consensus forecasts. In particular, communication related to the economy has a much stronger impact on both inflation and growth forecasts during the period starting with the Great Financial Crisis and ending before the quantitative easing phase in June 2014. After the ECB started to deploy its asset purchase program, money growth forecast revisions were associated strongly with the bank's communicated assessments of the course of the economy. Whereas communication on money overall has a positive impact on money growth forecasts (see Table 3.7), the relationship was reversed during the Great Financial Crisis and the euro crisis period. In the same period, the actual increase in assets and securities for monetary policy purposes led to a decline in money growth forecasts. Both effects appear to have been temporary, as the counterintuitive relationships have disappeared with the onset of quantitative easing.

	Change in inflation expectations			Change in GDP growth expectations			Change in M3 growth expectations		
	Full	Gen-to-spec.	SUR est.	Full	Gen-to-spec.	SUR est.	Full	Gen-to-spec.	SUR est.
Expected inflation in 12 months	-0.0676***	-0.0689***	-0.0675***	-0.144***	-0.141***	-0.140***	-0.0611		
(at previous consensus release)	(-3.108)	(-3.038)	(-4.298)	(-5.085)	(-4.986)	(-6.001)	(-0.852)		
Expected GDP growth in 12 months	-0.0173*			-0.0693***	-0.0660***	-0.0604***	-0.0110		
(at previous consensus release)	(-1.838)			(-3.554)	(-3.670)	(-4.987)	(-0.308)		
Expected M3 growth in 12 months	0.00510			0.00785			-0.0371*	-0.0382*	-0.0408**
(at previous consensus release)	(0.810)			(0.772)			(-1.663)	(-1.950)	(-2.441)
ECB Communication – price stability	0.155***	0.144***	0.144***	0.108**	0.121***	0.114***	0.0700		
(available at forecast release)	(4.559)	(4.291)	(5.029)	(2.449)	(2.835)	(2.783)	(0.573)		
ECB Communication – economy	0.0837***	0.0734***	0.0702***	0.268***	0.271***	0.256***	0.181*		
(available at forecast release)	(2.681)	(2.793)	(2.651)	(6.273)	(6.590)	(6.207)	(1.840)		
ECB Communication – money	-0.00216			-0.0347			0.347*	0.343**	0.364**
(available at forecast release)	(-0.0450)			(-0.523)			(1.824)	(1.993)	(2.570)
Main Refinancing Rate	0.221***	0.188**	0.188***	0.508***	0.525***	0.512***	0.0939		
(change between professional forecasts)	(3.056)	(2.582)	(3.808)	(5.094)	(6.005)	(7.226)	(0.428)		
Surprise in MRR (based on 3m Euribor futures)	0.231			-0.254			-1.205*		
(change since day before meeting)	(1.157)			(-0.670)			(-1.662)		
Assets & Securities for monetary policy purposes	-0.000700*	-0.000754**	-0.000697**	-0.00266***	-0.00267***	-0.00258***	-0.00188		
(change between professional forecasts)	(-1.826)	(-1.996)	(-2.043)	(-4.750)	(-4.741)	(-5.173)	(-1.574)		
Realised inflation rate	0.105***	0.107***	0.101***	0.0462			-0.0648		
(change between professional forecasts)	(3.483)	(3.625)	(3.840)	(0.876)			(-0.649)		
Realised annual GDP growth	0.0152			0.0732**	0.0753**	0.0737***	-0.0568		
(change between professional forecasts)	(0.851)			(2.177)	(2.362)	(3.507)	(-1.011)		
Realised annual M3 growth	0.00778			0.0114			0.169***	0.145**	0.141***
(change between professional forecasts)	(0.505)			(0.452)			(2.718)	(2.482)	(2.695)
Constant	0.0966**	0.104**	0.101***	0.272***	0.294***	0.287***	0.225*	0.128	0.137*
	(2.055)	(2.552)	(3.545)	(5.157)	(5.540)	(6.907)	(1.729)	(1.487)	(1.834)
Observations	188	188	188	188	188	188	188	188	188
Adjusted R-squared	0.370	0.375	0.394	0.478	0.483	0.502	0.107	0.108	0.122

Table 3.7. Monetary policy and communication impact on professional forecasters' expectations. The sample covers the May 2003 – December 2018 period. The OLS results (in the columns labeled "Full" and "Gen-to-spec.") report robust t-statistics in parentheses.
***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

	Change inflation exp.		Change growth exp.		Change gr.M3 exp.	
	Full	Gen-to-spec.	Full	Gen-to-spec.	Full	Gen-to-spec.
Expected inflation in 12 months	-0.0875***	-0.0906***	-0.105***	-0.0825***	-0.00227	
(at previous release)	(-3.112)	(-4.055)	(-3.072)	(-3.075)	(-0.0277)	
Expected GDP growth in 12 months	-0.0147	-0.0184*	-0.0784***	-0.0748***	0.0142	
(at previous release)	(-1.359)	(-1.813)	(-3.191)	(-3.592)	(0.257)	
Expected M3 growth in 12 months	-0.0110		-0.00873		-0.127***	-0.104***
(at previous release)	(-1.119)		(-0.469)		(-3.320)	(-3.591)
ECB Communication - price stability	0.321***	0.175***	0.0312		0.197	
(available at forecast release)	(3.689)	(5.093)	(0.324)		(0.693)	
Dummy ECB phase 2008:9-2014:5	-0.170**		0.0342		-0.325	
* ECB communication - price stability	(-2.162)		(0.371)		(-1.088)	
Dummy ECB phase 2014:6-2018:12	-0.187**		0.0428		0.00276	
* ECB communication - price stability	(-2.119)		(0.469)		(0.00941)	
ECB Communication - economy	-0.0394		0.0823*	0.0791***	-0.0601	
(available at forecast release)	(-1.090)		(1.706)	(2.825)	(-0.530)	
Dummy ECB phase 2008:9-2014:5	0.188***	0.192***	0.384***	0.382***	-0.00671	
* ECB communication - economy	(3.024)	(3.541)	(3.331)	(3.526)	(-0.0270)	
Dummy ECB phase 2014:6-2018:12	0.0622		0.0208		0.656***	0.609***
* ECB communication - economy	(0.833)		(0.283)		(2.713)	(3.371)
ECB Communication - money	-0.00206		0.0840		0.391***	0.497***
(available at forecast release)	(-0.0420)		(1.059)		(2.657)	(2.876)
Dummy ECB phase 2008:9-2014:5	-0.112		-0.402***	-0.314***	-0.657*	-0.754**
* ECB communication - money	(-1.418)		(-3.034)	(-2.725)	(-1.656)	(-2.168)
Dummy ECB phase 2014:6-2018:12	-0.200		-0.285*	-0.185**	0.363	
* ECB communication - money	(-1.162)		(-1.854)	(-2.030)	(0.500)	
Main Refinancing Rate	0.0930	0.186***	0.196**	0.161*	0.132	
(change between professional forecasts)	(1.371)	(3.015)	(2.258)	(1.929)	(0.510)	
Dummy ECB phase 2008:9-2014:5	0.162		0.347**	0.443***	-0.207	
* change main refinancing rate	(1.492)		(1.988)	(3.202)	(-0.485)	
Dummy ECB phase 2014:6-2018:12	-0.225*	-0.240***	-0.119		0.760**	0.992***
* change main refinancing rate	(-1.888)	(-3.315)	(-1.032)		(2.095)	(4.804)
Surprise in MRR (based on 3m Euribor futures)	0.164		0.369		-0.187	
(change since day before meeting)	(0.949)		(1.280)		(-0.277)	
Dummy ECB phase 2008:9-2014:5	0.230		-0.761		-1.043	
* change implied future MRR	(0.798)		(-1.388)		(-0.832)	
Dummy ECB phase 2014:6-2018:12	-1.036		-0.108		-2.191	
* change implied future MRR	(-1.079)		(-0.201)		(-0.814)	
Assets & Securities for monetary policy purposes,	0.00130*		-9.81e-05		-0.00265	-0.00361**
2014:6-2018:12 (change between prof. forecasts)	(1.818)		(-0.162)		(-1.153)	(-2.448)
Dummy ECB phase 2008:9-2014:5	-0.00183		-0.00566***	-0.00583***	-0.00225	
* change monetary policy assets	(-1.500)		(-3.876)	(-3.958)	(-0.750)	
Realised inflation rate	0.0655**	0.0819***	-0.00130		-0.0295	
(change between professional forecasts)	(2.379)	(3.050)	(-0.0269)		(-0.275)	
Realised annual GDP growth	0.0161		0.0822***	0.0852***	-0.0346	
(change between professional forecasts)	(0.974)		(2.733)	(2.887)	(-0.601)	
Realised annual M3 growth	0.0157		0.0214		0.163**	0.157**
(change between professional forecasts)	(0.982)		(0.896)		(2.605)	(2.518)
Constant	0.263***	0.208***	0.355***	0.275***	0.692***	0.549***
	(3.117)	(5.232)	(2.790)	(4.710)	(2.842)	(3.429)
Dummy ECB phase 3	-0.0836**	-0.0660***	-0.0985*	-0.0935**	-0.294**	-0.228***
	(-2.237)	(-3.109)	(-1.797)	(-2.513)	(-2.223)	(-2.806)
Dummy ECB phase 4	-0.127***	-0.0857***	-0.0687	-0.0485**	-0.353**	-0.187*
	(-2.728)	(-3.429)	(-1.226)	(-2.108)	(-2.223)	(-1.715)
Observations	188	188	188	188	188	188
Adjusted R-squared	0.463	0.447	0.572	0.585	0.151	0.188
F-test interaction terms price stability zero	0.082		0.896		0.259	
F-test interaction terms economy zero	0.012		0.004		0.026	
F-test interaction terms money zero	0.239		0.007	0.003	0.210	
F-test interaction terms main ref.rate zero	0.023		0.041		0.041	
F-test interaction terms future MRR zero	0.381		0.378		0.556	
F-test interaction terms mon.pol. assets zero	0.136		0.000		0.454	

Table 3.8. Impact across different policy phases. Notes: The sample covers the May 2003 – December 2018 period. The OLS results (in the columns labeled “Full” and “Gen-to-spec.”) report robust t-statistics in parentheses.
***p $<$ 0.01, **p $<$ 0.05, *p $<$ 0.1.

3.5 Conclusion

In this paper, we have studied one particular channel through which the European Central Bank (ECB) communicates with the external world: the press releases following ECB Governing Council meetings at which monetary policy decisions are made. Those press releases have been coded at the statement level and allow us to analyze the tones the ECB adopts on topics related to price stability, the real economy and monetary phenomena. Arguably, those three topics capture the monetary policy mandate of the ECB.

We have first looked into the impact those messages have on the financial market. For that we use the shadow rate as developed and published by Krippner (2013a,b), which reflects the assessment of financial market participants of the ECB's overall monetary policy stance. By comparing the shadow rate the day before and the day after the ECB's council meeting, we are confident that our results are driven by the actions of the ECB. Of the three topics identified and analyzed, we find a strong and robust impact of communication regarding the real economy and price stability on the shadow rate. Inflation expectations as measured by interest-rate swaps also appear to be affected by ECB communication on monetary developments. None of the established relationships appears to have changed significantly over time. Hence, the short-run reactions of financial market participants in setting interest rates or forming inflation expectations after a monetary policy decision and its communication by the ECB has remained remarkably stable over time.

In a next step, we have analyzed how professional forecasters, on average, change their inflation, GDP and money growth outlooks for the upcoming twelve months using data from Consensus Economics, Inc. We use the change in those consensus monthly forecasts, i.e., between consensus forecast releases. Given that longer time window, we can be less certain that we are indeed estimating a truly causal impact. We do, however, find that soft(er) information revealed by the press releases generally are related to consensus forecast adjustments. Overall, the information contained in that communication device appears to outperform actual monetary policy adjustments as measured by the change in the main refinancing rate and the assets and securities on the ECB's balance sheet that are used for monetary policy purposes. Hence, again words seem to matter more than deeds.

In contrast to what we found for the shadow rate, professional forecasters did appear to have changed the ways in which they have interpreted the communication and decisions of the ECB. In particular, during the Great Financial Crisis and the euro area crisis, their behavior was different. To quite some extent, the relationship appears to have normalized again during the quantitative easing phase that started June 2014.

3.A Appendix

3.A.1 Tables

	Shadow Rate		Nominal effective exchange rate		Inflation expectations (swaps)	
	Gen-to-Spec	Gen-to-Spec	Gen-to-Spec	Gen-to-Spec	Gen-to-Spec	Gen-to-Spec
ECB communication – price stability	0.123**	0.139**	-0.00234		0.0221	
	(2.155)	(2.361)	(-0.0325)		(0.283)	
ECB communication – economy	0.139**	0.151**	0.0562		0.108	
	(2.153)	(2.369)	(0.865)		(1.273)	
ECB communication – money	0.0732		0.0791		0.219***	0.222***
	(1.242)		(1.268)		(3.059)	(3.183)
Decided change in main refinancing rate			-0.148***	-0.119**		
(change since day before meeting)			(-2.732)	(-2.371)		
Surprise in MRR (based on euribor futures)	0.581***	0.580***	0.435***	0.440***		0.186*
(change since day before meeting)	(8.560)	(8.445)	(6.170)	(6.412)		(1.727)
Surprise in US Jobless Claims					-0.202**	-0.164*
					(-2.049)	(-1.864)
Expected inflation in 12 months						0.166*
(change since previous meeting)						(1.839)
Expected GDP growth in 12 months	0.221***	0.212***				
(change since previous meeting)	(4.294)	(4.072)				
Expected M3 growth in 12 months			0.181**	0.205**	-0.154**	-0.147**
(change since previous meeting)			(2.213)	(2.598)	(-2.283)	(-2.234)
Observations	168	168	168	168	155	155
Adjusted R-squared	0.415	0.414	0.227	0.232	0.082	0.128

Table 3.9. Standardized coefficients of selected results presented in Table 3.3 and Table 3.5. Notes: The sample covers the period May 2003 – December 2018. Standardized (or beta) coefficients are shown.
*** p<0.01, ** p<0.05, * p<0.1

	Change in inflation expectations			Change in GDP growth expectations			Change in M3 growth expectations		
	Full	Gen-to-spec.	SUR est.	Full	Gen-to-spec.	SUR est.	Full	Gen-to-spec.	SUR est.
Expected inflation in 12 months	-0.346***	-0.353***	-0.346***	-0.460***	-0.450***	-0.446***	-0.103		
(at previous consensus release)	(-3.108)	(-3.038)	(-4.298)	(-5.085)	(-4.986)	(-6.001)	(-0.852)		
Expected GDP growth in 12 months	-0.149*			-0.372***	-0.355***	-0.324***	-0.0312		
(at previous consensus release)	(-1.838)			(-3.554)	(-3.670)	(-4.987)	(-0.308)		
Expected M3 growth in 12 months	0.0705			0.0675			-0.169*	-0.174*	-0.185**
(at previous consensus release)	(0.810)			(0.772)			(-1.663)	(-1.950)	(-2.441)
ECB Communication – price stability	0.407***	0.379***	0.380***	0.176**	0.199***	0.187***	0.0607		
(available at forecast release)	(4.559)	(4.291)	(5.029)	(2.449)	(2.835)	(2.783)	(0.573)		
ECB Communication – economy	0.205***	0.180***	0.172***	0.408***	0.412***	0.390***	0.146*		
(available at forecast release)	(2.681)	(2.793)	(2.651)	(6.273)	(6.590)	(6.207)	(1.840)		
ECB Communication – money	-0.00378			-0.0378			0.200*	0.197**	0.210**
(available at forecast release)	(-0.0450)			(-0.523)			(1.824)	(1.993)	(2.570)
Main Refinancing Rate	0.279***	0.238**	0.237***	0.398***	0.412***	0.402***	0.0390		
(change between professional forecasts)	(3.056)	(2.582)	(3.808)	(5.094)	(6.005)	(7.226)	(0.428)		
Surprise in MRR (based on euribor futures)	0.0737			-0.0503			-0.126*		
(change since day before meeting)	(1.157)			(-0.670)			(-1.662)		
Assets & Securities for monetary policy purposes	-0.159*	-0.171**	-0.158**	-0.375***	-0.377***	-0.365***	-0.141		
(change between professional forecasts)	(-1.826)	(-1.996)	(-2.043)	(-4.750)	(-4.741)	(-5.173)	(-1.574)		
Realised inflation rate	0.241***	0.247***	0.232***	0.0662			-0.0491		
(change between professional forecasts)	(3.483)	(3.625)	(3.840)	(0.876)			(-0.649)		
Realised annual GDP growth	0.0602			0.181**	0.186**	0.182***	-0.0743		
(change between professional forecasts)	(0.851)			(2.177)	(2.362)	(3.507)	(-1.011)		
Realised annual M3 growth	0.0339			0.0309			0.242***	0.208**	0.203***
(change between professional forecasts)	(0.505)			(0.452)			(2.718)	(2.482)	(2.695)
Observations	188	188	188	188	188	188	188	188	188
Adjusted R-squared	0.370	0.375	0.394	0.478	0.483	0.502	0.107	0.108	0.122

Table 3.10. Standardized coefficients of results presented in Table 3.7. Notes: The sample covers the period May 2003 – December 2018. The OLS results (in the columns labeled "Full" and "Gen-to-spec.") report robust t-statistics in parentheses. Standardized (or beta) coefficients are shown.
*** p<0.01, ** p<0.05, * p<0.1.

3.A.2 Figures

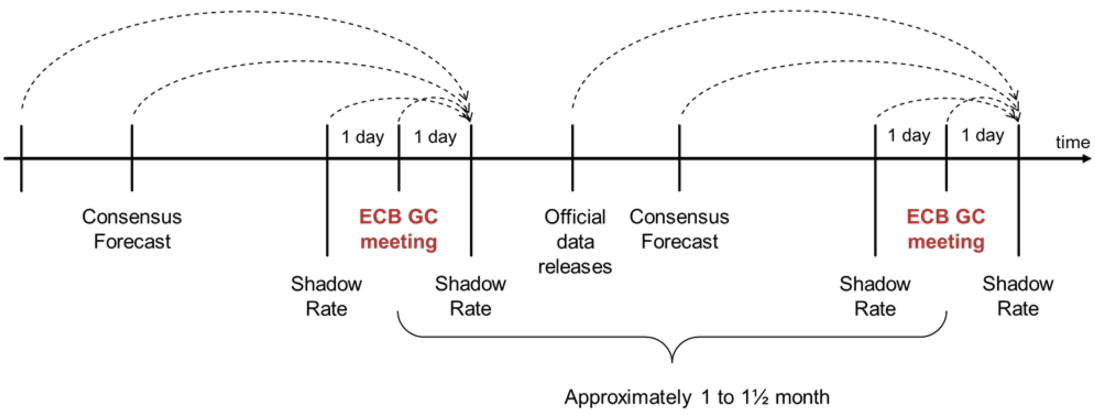


Figure 3.4: Data Timeline for Results in Tables 3.1 – 3.5 and 3.9.

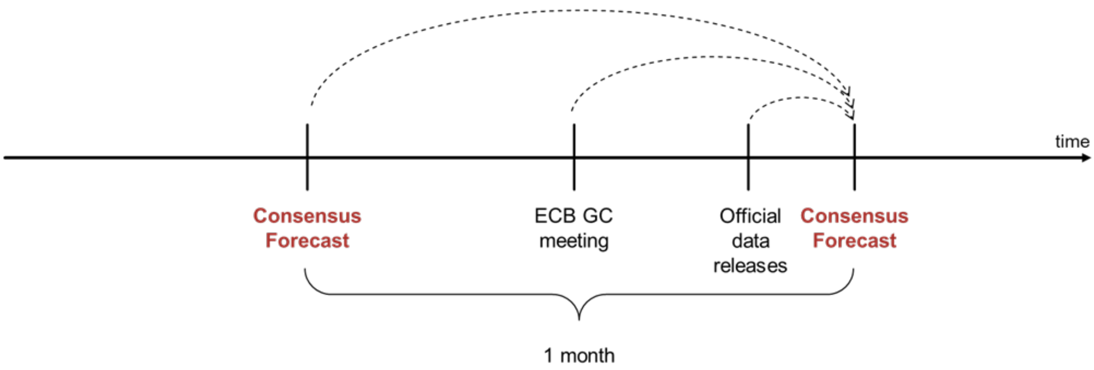


Figure 3.5: Data Timeline for Results in Tables 3.6 – 3.8 and 3.10.

Chapter 4

Mental Health Effects of Social Distancing in Switzerland[1]

4.1 Introduction

The outbreak of the COVID-19 pandemic at the beginning of 2020 resulted in drastic measures to bring down infection rates. A significant part of public health policies revolved around social distancing and, more generally, reduction in mobility (Ayouni et al. 2021). Although important to stop the spread, these measures led to a decrease in economic activity and a pronounced increase in unemployment (Deb et al. 2022). In addition to economic uncertainty and decreased profitability in most sectors, it is important to understand how the COVID-19 pandemic and related measures affect public mental health.

Most variables that capture mental health are based on individual survey data, which are prone to measurement error (Bound et al., 2001). To reduce this measurement error, we use administrative data from the most prominent Swiss helpline, *Offering a Helping Hand*, to examine the mental health effects of the COVID-19 pandemic and associated policies. *Offering a Helping Hand* is a phone service that provides free counseling. To ensure a complete database, we merge the manually

[1]This chapter is joint work with Stefan Pichler.

recorded call listings with administrative data from the local telephone network operator. The central aim of this paper is to find out if the number and duration of calls to the helpline increased after the outbreak of the pandemic. To achieve this goal, we follow an event study regression framework.

Our paper contributes to the rapidly growing literature on the effects of the pandemic on mental health. Using mental health surveys, Costa-Font et al. (2022), Mendez-Lopez et al. (2022), Muresan et al. (2022), Serrano-Alarcón et al. (2022), Burdett et al. (2021), Richter et al. (2021), and Banks and Xu (2020) find worse mental conditions after the outbreak.[2] Online search behavior reveals a similar pattern, as searches related to anxiety (Fetzer et al. 2021, Silverio-Murillo et al. 2021) and sadness (Brodeur et al. 2021) are reported to increase during lockdowns. Fortunately, the strain that the pandemic put on mental health appears to have not passed through suicide rates, which are reported to have declined during the first wave of COVID-19 (Pirkis et al. 2021, Tanaka and Okamoto 2021).

Another tool to measure mental health is helpline call data. These data are available at high frequency and can be used as a proxy for help-seeking and mental health in the general population. The suitability for assessing mental health has spurred research in this area since the outbreak (Monreal-Bartolomé et al. 2022, Batchelor et al. 2021, Zalsman et al. 2021,Turkington et al. 2020, Halford et al. 2020, Armbruster and Klotzbücher 2020).[3] Assembling helpline data from 19 countries, Brülhart et al. (2021) (from now on BKLR), use data from manually recorded call listings to show that call volumes increased significantly during the first wave. Compared to before the pandemic, the number of calls is reported to have increased by 35% at the six-week peak.

Our study differs from this literature in two important dimensions. First, all the papers except Batchelor et al. (2021) and Bullinger et al. (2021) rely on manually recorded data entries. In our paper, we make use of administrative data from the

[2]This is in line with worse mental health in the aftermath of the financial crisis as documented by McInerney et al. (2013) and Phillips and Nugent (2014). On the contrary, Baird et al. (2013) finds an increase in mental health after a positive income shock.

[3]In related studies Erten et al. (2022), Leslie and Wilson 2020, and Bullinger et al. (2021) find increases in domestic violence related to the pandemic and social distancing.

network provider. Apart from being a comprehensive data source, it also allows us to identify repeat callers, as we have a unique hashed caller identifier in our data set. Moreover, we can identify the unmet need for help-seeking as our data also contain calls that were not answered. This is important since unmet need is found to be a determinant of future health status (Weathers and Stegman 2012, Gibson et al. 2019). Comparing our results to Brülhart and Lalive (2020) (henceforth BL), we show that administrative data matters. We can replicate their insignificant results for Switzerland using data from the helpline. However, once we use data from the network provider, we find pronounced and highly significant increases in call contacts and duration.

Second, our data allow us to analyze both the first COVID-19 wave in February 2020 and the second wave in the late fall of 2020. It is interesting to analyze the second wave for Switzerland because the social distancing policies were much less restrictive during the second wave (both compared to the first wave in Switzerland and compared to the second wave in other countries in western Europe). This allows us to compare (i) the mental health effects at the beginning of the pandemic combined with strict social distancing measures with (ii) the mental health effects in the second wave with less restrictive policies. Our results show increased frequency and duration of calls during the first wave. On the other hand, we do not find significant effects for the second wave. This provides suggestive evidence that strict social distancing measures might lead to worse mental health.

The closest match to our study is BL, which published a preliminary analysis based on manually recorded call listings of Switzerland's most popular helpline. Since the analysis was carried out only two months after the first measures were taken by the government, it only includes parts of the first COVID-19 wave. They find no significant effects on the total number of calls after the outbreak, which is in contrast to the findings in BKLR based on a more comprehensive set of international helplines. However, BL do not use administrative data from the network operator, nor analyze the duration of calls or whether calls are answered by the helpline.

Using difference-in-differences and pre-post study methods on the complete call database from the network provider, our results reconcile the Swiss case with the findings in BKLR. For the first wave, we find a significant increase in the number of calls made to the helpline, both in absolute terms and relative to the corresponding control period in 2019. While total call duration also increased after the outbreak, this effect would likely have been much stronger if there had been more supply capacity, as we find a pronounced increase of calls that were both either not answered right away or never answered by the helpline (unmet need).

Moreover, we find the increase in calls to be driven by people above the age of 65, but there is also a firm reaction from individuals below that threshold.[4] Unlike the literature based on cross-sectional data, the beginning of the pandemic caused men to seek more help compared to female callers in Switzerland. Finally, we find small and sometimes negative effects during the second wave. A potential explanation is that Switzerland enforced a relatively lax social distancing regime in this period. However, given that the policy measures were relatively homogeneous in all Swiss cantons, we cannot test whether stricter measures would have led to an increase in helpline calls.

The remainder of this paper proceeds as follows. Section 2 provides background information on the chronology of the COVID-19 outbreak in Switzerland and on the helpline *Offering a Helping Hand*. Section 3 summarizes the data from the counseling service and the telephone network operator. The econometric approach is described in Section 4 while Section 5 presents the empirical findings. Section 6 presents the results of several robustness checks and Section 7 concludes.

4.2 Background

This section provides a short overview of the evolution of COVID-19 in Switzerland and introduces the helpline *Offering a Helping Hand*.

[4]There is a different helpline for children and juveniles not analyzed in this paper. This is due to the lack of data from that helpline and a potential concern related to underreporting for children and adolescents, as documented in (Baron et al., 2020).

4.2.1 COVID-19 in Switzerland

Being close to Italy – an early COVID-19 hotspot in Europe – Switzerland registered the first COVID-19 case on February 25, 2020. Figure 4.1 shows the evolution of new infections as a seven day moving average along with the stringency index of the Oxford COVID-19 government response tracker for Switzerland (Ritchie et al. 2020). The first wave is characterized by a (in hindsight) relatively mild infection curve but a strong government response. The sudden jump in the stringency index can be explained by the introduction of a nationwide lockdown, with nonessential stores, schools, and recreational facilities being forced to shut down and nonurgent medical procedures halted.

The second wave, starting at the beginning of October, features an unprecedentedly large number of confirmed cases, while the reaction of the stringency index is relatively muted when compared to earlier in the year. In fact, the federal government abstained from introducing a second lockdown, but restricted the number of people who could meet in private and closed nightlife activities. This is very different from other parts of Europe where the government response was much stronger during the second wave. In Switzerland, apart from closing restaurants and recreational facilities again in December, measures were kept almost unchanged until March 31, 2021, which constitutes the end of our observation period.

4.2.2 Helpline *Offering a Helping Hand*

The Swiss telephone helpline, *Offering a Helping Hand* ("Dargebotene Hand" in German - henceforth DH), is a well-known nationwide service that is funded through charitable donations. The helpline is available 24-7 and provides free counseling by trained volunteers, adhering to the standards of the International Federation of Telephone Emergency Services (IFTES). Although calls are anonymous, helpline operators note their best guess of the gender and age category of the caller. The helpline is organized as an association of twelve regional call centers spread throughout Switzerland, as shown in Figure 4.2. Callers are automatically connected to a fixed helpline center based on their location, taking into account their respective

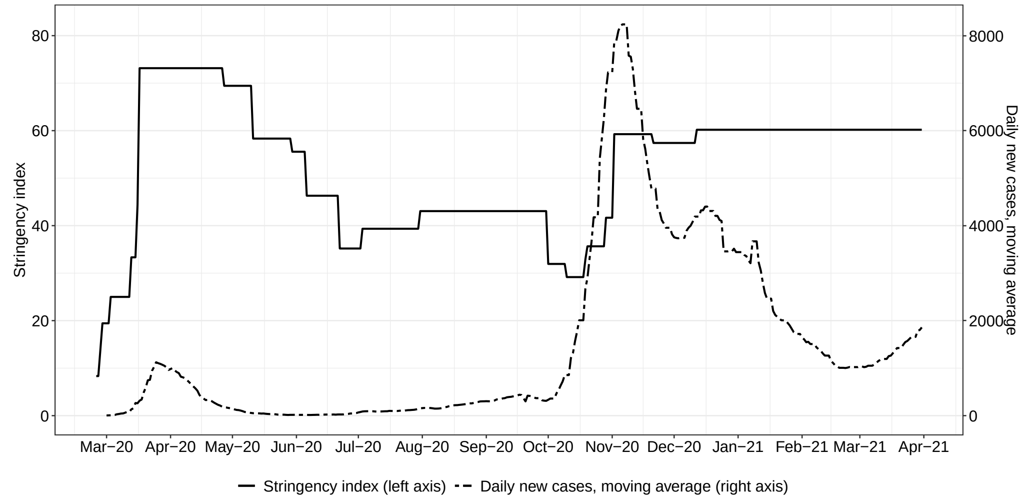


Figure 4.1: COVID-19 infections and stringency index. Notes: This figure shows daily COVID-19 infections and the stringency index for Switzerland. Infections are plotted as seven day moving averages (right axis, dotted line) in order to reduce variability. The stringency index (left axis, straight line) is taken from the Oxford COVID-19 government response tracker and ranges from 0 to 100 with the latter indicating the most restrictive policy regime. Source: Ritchie et al. (2020).

language regions. If there is no capacity, calls are sometimes forwarded to another helpline center (see the discussion in the next section).

4.3 Data

Here we present more information concerning our data sources and outcome and explanatory variables.

4.3.1 Data Sources

Following BL, we obtained detailed call-level data from the DH helpline which was manually recorded by the volunteers answering the calls. The observation period starts in the beginning of 2015 and ends in March 2021. The data set consists of notes from the volunteer answering the phone call, recording the time of the call, as well as approximate duration, age, and gender. However, these data are likely to

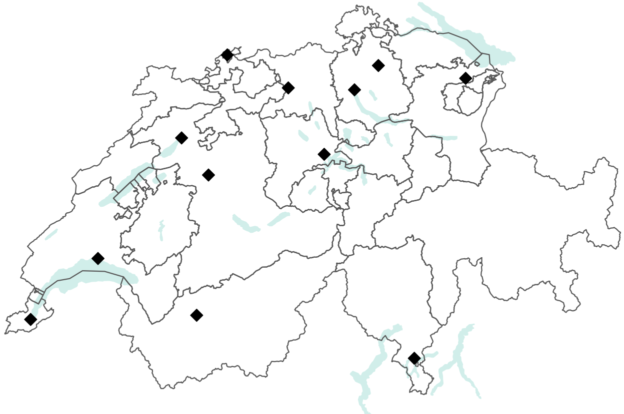


Figure 4.2: Regional helpline centers. Notes: This figure depicts the Swiss map together with all regional helpline centers. For each of the twelve centers, a black diamond shows the geographical location. Source: DH.

contain some measurement error, and only 9 out of 12 call centers provided these data. Finally, some call centers started collecting data only later in the sample period. From January 2017 to March 2021 (our estimation period), the helpline data consist of a total of 806,739 call entries.

To obtain a more reliable data set, we contacted the phone network operator that is in charge of the helpline and received the complete records for the 12 call centers, including a hashed calling number. These data were obtained for the same observation period as in the helpline database (January 2015 - March 2021). Comparing the summary table 4.12 with the helpline table reveals large differences in terms of call coverage. There are a total of 1,632,925 calls listed in the data set between January 2017 and March 2021, i.e., more than twice as much as in the manually collected call history. Judging from the distribution of call duration in both data sources (Table 4.13), volunteers appear to systematically underreport calls with a low duration.

We work with both data sets, as they provide complementary information. The advantage of network provider data is better coverage, while helpline listings contain demographic information on the caller.

4.3.2 Outcome Variables

We use the number and duration of calls as outcome variables. Figure 4.3 shows a scatter plot with weekly call number[5] and total duration for the entire observation period for the network provider database. For completeness, the figure also shows that the average duration is decreasing when we divide the total duration by the number of calls. Since the average duration is simply the ratio of the number of calls and duration, we focus only on the latter two in our regressions. Although most of our computations are done using daily data, but we also aggregate outcome variables to a weekly frequency in the robustness section.

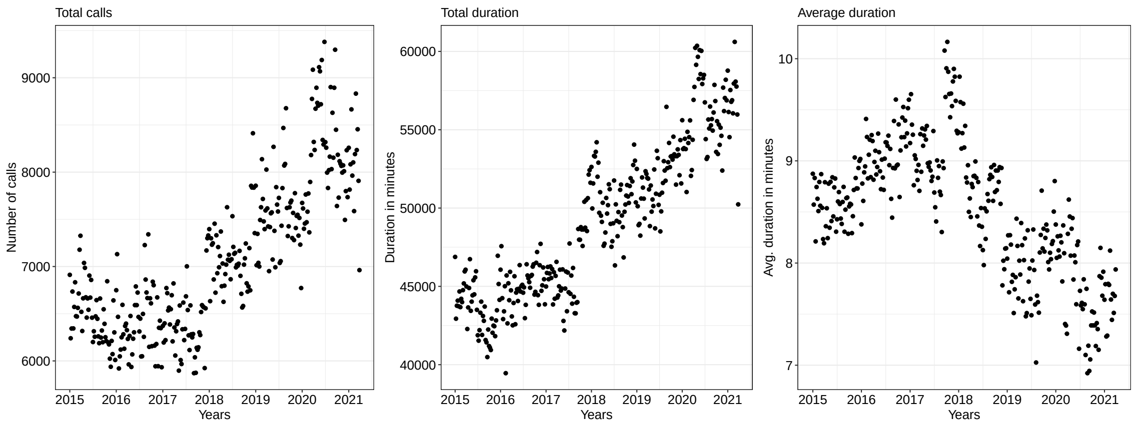


Figure 4.3: Nationwide call number and duration. Notes: For the entire observation period, this figure shows scatter plots for total call number (left), total duration (middle) and average duration per call (right). To gain a better overview, each dot represents either the weekly sum (total calls and duration) or the weekly ratio (average duration) instead of daily values. Source: Network provider.

4.3.3 Control Variables

Although the call record from the helpline is rather incomplete, we can use it to add information about age and gender to the call listings from the network provider. The merging is performed for each regional center using the exact call time and

[5] All our analyses include calls that were unanswered. In a separate analysis in Section 4.5.3, we separate answered and unanswered calls.

duration information available for both data sets. Using the calling number from the network data, we then extrapolate the age and gender information to calls from the same identifier where merges were unsuccessful.[6] In other words, we assume that calls from the same phone number are made by the same person. Although there is room for matching mistakes, more than half of all calls come from mobile phones (53%), which arguably are not used by an entire household.[7]

Another potential problem for matching is that the network operator only records where the call was first placed, i.e., the closest helpline center from the calling position. However, if there are no spare capacities available, the call can be automatically forwarded to another regional center. For such calls, age and gender information could not be merged since the regional center entry diverges between the helpline and network provider listings. We address this issue using detailed call data from the telecommunication software provider, which supplies and maintains the telephone software system for nine regional centers. This allows us to identify forwarded calls to correct the network provider information on where the call actually was answered. Calls are only forwarded between five regional branches equipped with this system; hence, there is no need for correction for calls first answered by the remaining helpline centers.[8]

4.3.4 Samples

To analyze the first wave we use data from the beginning of January 2017 to 20 July 2020. As discussed below, we use different subsamples depending on the model specification. For the second wave, we use data starting in June 2020 and ending on 14 March 2021.

---

[6] From 806,739 helpline observations we can successfully merge 641,356 to the network provider call listings (matching rate of 80%). This allows to characterize 1,465,325 network provider observations with information on gender and age (90% of all entries, see also Table 4.11).

[7] The fact that someone calls from a mobile phone is revealed by the prefix of the calling number. For 11% (26%) of calling numbers, we have more than one entry for gender (age group). In this case, the most frequent case as recorded by the call agents is used.

[8] Around 10% of all calls are forwarded in regions that forward calls to each other.

The sample for the main analysis is aggregated on the daily and regional level. In the robustness section, we additionally use the caller identifier to construct a sample on the individual level. In terms of frequency this second sample is at the weekly level. Figure 4.4 shows the distribution of phone calls per phone number for the entire sample period. As the figure shows, the helpline has received only one call from almost 40,000 phone numbers. Note that for some phone numbers, significantly more than 20 calls were received.

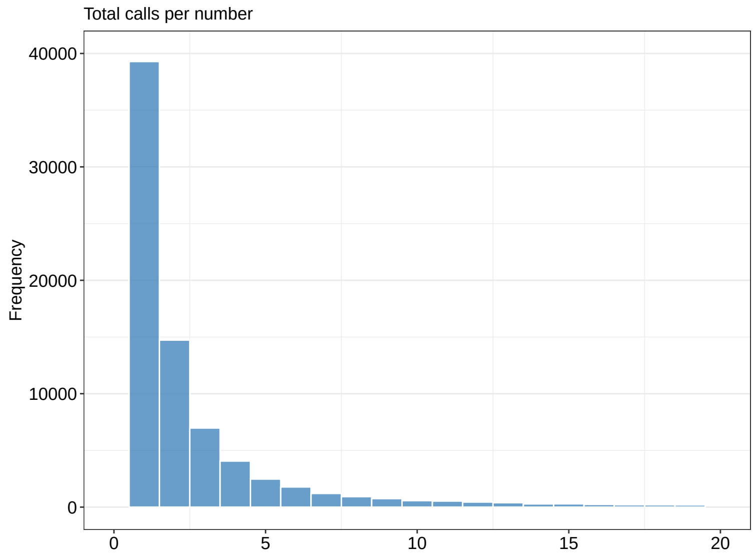


Figure 4.4: Distribution of number of calls per number and year. Notes: This figure shows the distribution of phone calls per phone number for the full sample period (Jan 2017 to March 2021). Note that some phone numbers have called more than 20 times, but are not on the histogram for visibility reasons. Source: Network provider.

4.4 Empirical Approach

This section explains the empirical methodology underlying our study. To assess how the COVID-19 pandemic affects public mental health in Switzerland, we apply an pre-post study design to the daily number and duration of helpline calls. Using our panel of daily call information for 12 regional centers, we estimate the following baseline model:

$$\ln(\text{Calls}_{r,t}) = \beta\text{Post}_t + \delta_d + \eta_r + \mu_w + \epsilon_{r,t} \tag{4.1}$$

where the dependent variable is the natural logarithm of the number or duration of calls to regional center r recorded on day t. The indicator variable Post_t is set equal to 1 for all days after which the stringency index a) turns positive for the first wave (February 25, also the first confirmed case), and b) increases again in fall 2020 for the second wave (October 19). The coefficient β therefore allows us to assess whether there has been a significant difference between before and after the outbreak of the first and second waves. Since the dependent variable is log transformed, it can be interpreted as the percentage deviation in daily calls. To prevent these coefficients from picking up the time trend present in the data, we include a weekly linear time trend μ_w to capture the long-term increase in the number and duration of calls.[9] The model also includes day-of-week fixed effects δ_d. To control for constant characteristics related to regional centers, we also add call center fixed effects η_r.

To capture the effect of the pandemic with more granularity we also estimate the following specification:

$$\ln(\text{Calls}_{r,t}) = \sum_{\tau=-16}^{-1} \beta^\tau \text{Week}_t^\tau + \sum_{\tau=1}^{22} \beta^\tau \text{Week}_t^\tau + \delta_d + \eta_r + \mu_w + \epsilon_{r,t} \tag{4.2}$$

where the base period ($\tau = 0$) for the two waves is again defined as the week the stringency index a) turns positive and b) increases in the fall of 2020. The indicator variable Week^τ is set 1 for all days of event week τ. Its coefficient β^τ thus allows us to assess the significance of call dynamics in week τ relative to the base week.

The third model we estimate is a difference-in-differences (DiD) model with 2019 as control and 2020 as treatment group. In other words, we compare the daily evolution of calls after the outbreak of the pandemic to the same days in the previous

[9] In an earlier version of the paper we omitted the linear time trend and received similar results, but the event study graphs showed a clear trend.

(pre-COVID) year:

$$\ln(\text{Calls}_{r,l,y}) = \beta(\text{Post}_l \times 2020_y) + \delta_d + \eta_r + 2020_y + \text{Post}_l + \epsilon_{r,l,y} \tag{4.3}$$

where the explained variable is the logarithm of the number or duration of calls to regional center r in year y on the day of the year l. The coefficient of interest is again β, which links the interaction of variable 2020_y (treatment indicator, which is one for 2020) and Post_l (day indicator) to the number of calls. While the trend is controlled for in models (4.1) and (4.2) by including a weekly linear time trend, this approach automatically controls for trends as 2020 is compared to 2019.

4.5 Results

The first part of this section, shows descriptive graphical evidence and results from regression analysis. The second part provides a heterogeneity analysis and investigates the question of capacity constraints.

4.5.1 Graphical Analysis

Figure 4.5 shows the development of the daily number, duration, and average duration of helpline contacts for the first wave (upper row, 4.5a) and the second wave (bottom row, 4.5b) in Switzerland. We compare the evolution in the pandemic year (in red) with the corresponding values on the same dates in 2019 (in blue). To facilitate comparison, solid trend lines are fitted using a local polynomial regression.

Looking at the trend line, the number of calls in the period before the outbreak of the first wave is similar to that of the previous year. In the following weeks, calls increase sharply and peak after around ten weeks. The evolution of call duration resembles an analogous behavior, increasing by 12% after ten weeks in 2020 while declining slightly in the control year 2019. In contrast, the average duration of helpline calls follow the same pattern in both years.

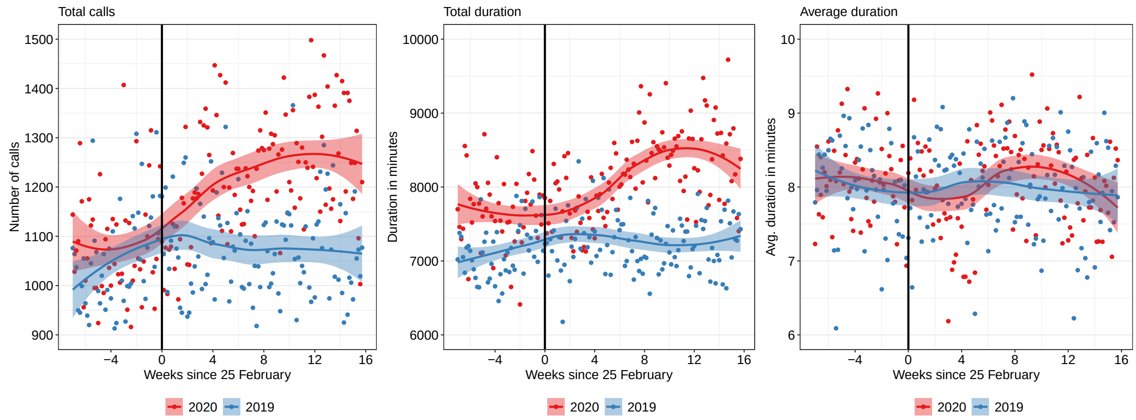
The second COVID-19 wave exhibits call dynamics quite different from the first one. While total duration exhibits a minor positive reaction after one month, this seems to be mainly driven by a catch-up in the average duration of phone calls. For the evolution of the number of call contacts there is no visible difference when compared to the year before. The trend line for daily calls and duration is uniformly higher in 2020 compared to 2019, while the opposite applies for average duration. This pattern corresponds to the positive (number and duration) or negative (average duration) time trend present in the data.

It is noteworthy, that this finding for the second wave is not in line with the evidence in BKLR. For Germany and France, the authors' evidence suggests that the volume of calls increases also during the second wave, especially for the case of France. This is remarkable given the high death rate in Switzerland. In fact, Switzerland had the highest number of confirmed deaths in relation to population among all three countries from November until mid-January. However, the government response was not as restrictive as in France or Germany, indicating that the policy reaction could have been more important for public mental health than actual mortality (see Appendix Figures 4.11 and 4.12).

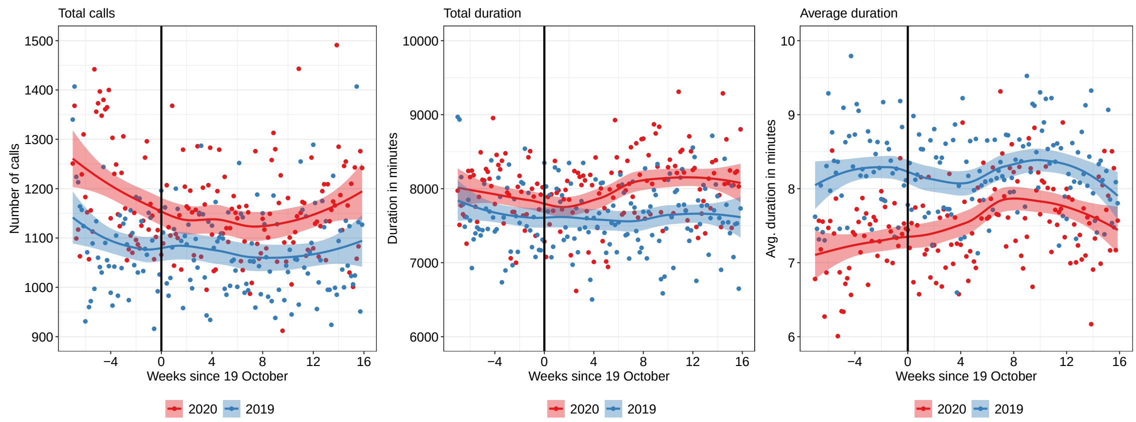
4.5.2 Main Results

In this section, we report regression results of models (4.1)-(4.3) for the number of contacts and the duration of calls as dependent variables. Since we include a linear weekly time trend in our routine, we restrict the sample to observations after January 2017, as suggested by the trend break seen in Appendix Figure 4.10. For the pre-post model (4.1) in the first wave, we provide estimates based on data from i) January 2017 to 20 weeks after the reference day, as well as ii) a symmetric time window of 20 weeks around the reference day. For the second wave, we provide estimates based on the symmetric window to exclude the first wave from the sample. Since the difference-in-differences model (4.3) compares the pandemic year 2020 to the previous year, the data ranges from January 1 to 20 weeks after the reference day for the first wave. The second wave is estimated using data from 20 weeks before the reference day to the end of December.

Table 4.1 shows the results of the pre-post model (4.1) and the DiD model (4.3) for the first wave. Looking at the effect the pandemic had on the number



(a) First COVID-19 wave (25 February 2020)



(b) Second COVID-19 wave (19 October 2020)

Figure 4.5: Daily call counts and (average) duration. Notes: For the first and the second wave, the scatter plot shows daily helpline calls and duration before and after the reference day for the years 2019 and 2020. The vertical black line depicts the reference day for the first (25 February 2020) and the second (19 October 2020) wave. The colored solid lines are fitted using local polynomial regression, while the surrounding area represents the 95% confidence intervals. The upper graphs range from January 11 to June 19 and refer to the first wave. The lower graphs cover the period from August 31 to February 7 including the time of the second wave of COVID-19 infections. Source: Network provider.

of helpline contacts (first 3 rows), the positive and significant coefficients confirm the descriptive graphical analysis. Relative to the base period, the number of calls is estimated to increase between 7.9 and 10.7% in the twenty weeks after the first confirmed COVID-19 case. The effect on total duration of daily calls is similar although less pronounced. Depending on the base period, total duration of daily phone calls increases by around 4.8 to 5.4% in the 20 weeks after the first wave started.

Dependent Variable:	ln(Calls)			ln(Duration)		
Estimated equation:	(1)	(1)	(3)	(1)	(1)	(3)
Coefficient						
$\hat{\beta}$	0.079***	0.106***	0.107***	0.054***	0.048***	0.048***
	(0.008)	(0.019)	(0.018)	(0.006)	(0.014)	(0.014)
Analysis period						
01 January 2017 - 20 July 2020	✓			✓		
08 October 2019 - 20 July 2020		✓			✓	
01 January - 20 July 2020/2019			✓			✓
Pretreatment mean	83.9	89	89.4	591.8	635.8	620.4
Observations	15,564	3,444	4,824	15,564	3,444	4,824
R^2	0.62	0.66	0.66	0.76	0.78	0.78

Table 4.1. Pre-post study: main findings for the first COVID-19 wave. Notes: The table shows estimates from pre-post regression model (4.1) and DiD model (4.3) for the first COVID-19 wave. $\hat{\beta}$ is the coefficient from the indicator variable Post_t, which is set zero before 25 February 2020 and one afterwards. Model (4.1) is estimated using data from a) 01 January 2017 to 20 July 2020 (i.e. ending 20 weeks after the reference day) as well as b) a symmetric time window of 20 weeks around the reference day. Model (4.3) uses data from the 01 January to 20 July in years 2020 and 2019. The first three rows show results for call count as dependent variable, while subsequent rows do so for call duration. Robust standard errors are given in parentheses. Significance codes: ***: 0.01, **: 0.05, *: 0.1. Source: Network provider.

Figure 4.6 shows the estimated weekly indicators characterized in model (4.2). Weekly contact volumes begin to increase shortly after the first confirmed case and peak three months after the outbreak (+22.5% compared to the base week). The increase in call activity is long-lasting and is highest between weeks 3 and 14 (on average + 19.5%), after which the volumes gradually decrease again to pre-pandemic levels. In line with the results from the pre-post study, total call duration increases significantly, although not as pronounced as the total number of calls. We observe positive effects mostly between weeks 7 and 17 with a peak call duration in week 9 (+12%).

Using complete network provider call data, we thus arrive at different conclusions than BL, who do not find an increase of call volumes in Switzerland during the first wave (above usual trend growth). In fact, estimating baseline models (4.1) and (4.2) with helpline data (as in BL) reproduces their main result. We find negative and insignificant values for total call counts in the pre-post study and no pattern in the estimated weekly indicators during their observation period (see Appendix Table 4.9 and Figure 4.9). The divergent results are arguably due to lower coverage of manual helpline records, also due to capacity constraints during the first wave.[10]

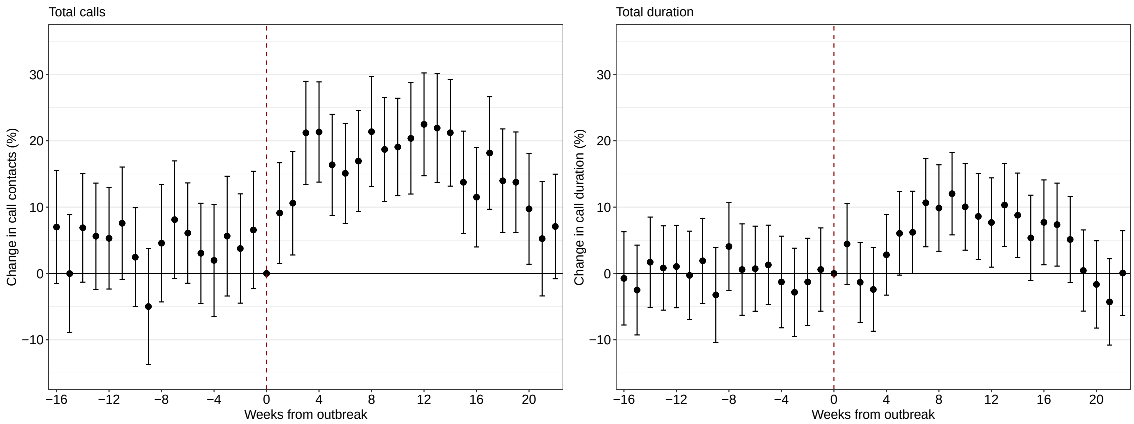


Figure 4.6: Event study: main findings for the first COVID-19 wave. Notes: The graph shows point estimates together with 95% confidence intervals from the weekly event study regression model (4.2) for the first COVID-19 wave. Week zero serves as base week and lasts from 25 February 2020 to 02 March 2020. All coefficients are estimated relative to the base week. Robust standard errors are used for calculating uncertainty measures. Source: Network provider.

The output from models (4.1)-(4.3) for the second COVID-19 wave support the previous graphical findings. The pandemic effects differ remarkably between first and second wave. Both the baseline and the DiD model tend to find either a slight decrease in the number of call contacts and duration or fail to reject the null hypothesis (Table 4.2). The estimated weekly indicators also remain largely insignificant

[10]For call duration (instead of volumes) we do find a significant effect of about the same magnitude as the results from the network provider database. However, BL only analyze call counts.

for both the number of contacts and total duration (Appendix Figure 4.8). Therefore, mental suffering related to the second COVID-19 wave is relatively contained in Switzerland.

Dependent Variable:	ln(Calls)		ln(Duration)	
Estimated equation:	(1)	(3)	(1)	(3)
Coefficient				
$\hat{\beta}$	-0.048***	-0.019	-0.002	-0.027**
	(0.018)	(0.016)	(0.014)	(0.012)
Analysis period				
01 June 2020 - 14 March 2021	✓		✓	
01 June - 31 December 2020/2019		✓		✓
Pretreatment mean	86.8	95.4	605.5	643.7
Observations	3,444	5,136	3,444	5,136
R^2	0.65	0.65	0.79	0.77

Table 4.2. Pre-post study: main findings for the second COVID-19 wave. Notes: The table shows estimates from pre-post regression model (4.1) and DiD model (4.3) for the second COVID-19 wave. $\hat{\beta}$ is the coefficient from the indicator variable Post_t, which is set zero before 19 October 2020 and one afterwards. Model (4.1) is estimated using a symmetric time window of 20 weeks around the reference day. Model (4.3) uses data from 01 June to 31 December in years 2020 and 2019. The first two rows show results for call count as dependent variable, while subsequent rows do so for call duration. Robust standard errors are given in parentheses. Significance codes: ***: 0.01, **: 0.05, *: 0.1. Source: Network provider.

Confirming the earlier graphical analysis (but unlike the evidence in BKLR for Germany and France), public mental health measured through helpline calls appears to remain stable in Switzerland during the second wave. Since the stringency indices were relatively homogeneous in Swiss cantons (see Appendix Figure 4.13), we cannot test whether stricter measures would have led to increased helpline call activity. As a result, we focus on the first wave in the remainder of the paper. Since for the first wave results for the baseline pre-post and the DiD model are almost identical, we report only the former model in what follows.

Dependent Variable:	ln(Answered)		ln(Unanswered)		ln(Never Answered)	
Coefficient						
$\hat{\beta}$	0.011	0.029*	0.299***	0.462***	0.411***	0.760***
	(0.008)	(0.017)	(0.029)	(0.070)	(0.055)	(0.131)
Analysis period						
01 January 2017 - 20 July 2020	✓		✓		✓	
08 October 2019 - 20 July 2020		✓		✓		✓
Pretreatment mean	69.5	77.8	14.3	11.1	2.7	2.4
Observations	15,564	3,444	15,564	3,444	15,564	3,444
R^2	0.70	0.72	0.29	0.37	0.15	0.20

Table 4.3. Pre-post study: capacity constraints during the first COVID-19 wave. Notes: The table shows estimates from pre-post regression model (4.1) for the first COVID-19 wave for answered and unanswered calls. $\hat{\beta}$ is the coefficient from the indicator variable Post_t, which is set zero before 25 February 2020 and one afterwards. The model is estimated using data from a) 01 January 2017 to 20 July 2020 (i.e. ending 20 weeks after the reference day) as well as b) a symmetric time window of 20 weeks around the reference day. The first two rows show results for the answered call count as dependent variable, while the subsequent two rows do so for unanswered calls. The final two rows show estimates for never answered calls, which is an adjusted version of unanswered calls. Robust standard errors are given in parentheses. Significance codes: ***: 0.01, **: 0.05, *: 0.1. Source: Network provider.

4.5.3 Capacity Constraints?

To find out whether there were capacity constraints during the first wave, we estimate models (4.1) and (4.2) separately on the number of answered and unanswered calls. Unanswered calls consist of all contacts that were not answered or found a busy line. In times of capacity constraints, the same person might try to call several times before finally receiving an answer, so we additionally calculate a measure of calls that have never been answered for each calling number, without double-counting subsequent calls for the same number. More precisely, this measure does not count unanswered calls that are followed by an answered call within 24 hours. To also prevent inflated counts from close succession of multiple call attempts, we additionally discard unanswered calls that are closer than one hour to each other. This allows us to see whether the unmet need has increased as a result of the pandemic.

Table 4.3 presents evidence that the Swiss helpline faces from pronounced capacity constraints, as unanswered calls increase between 30 and 46% depending on the sample used. Moreover, the subsample of calls that are never answered by the helpline increases between 41 and 76%. On the contrary, the call contacts that are answered seem to only increase slightly. The estimated weekly indicators of model (4.2) suggest that twelve weeks into the pandemic, the number of unanswered calls is 75% higher than in the base week (Figure 4.7). The number of never-answered calls doubles after seven weeks (+97%). While in many weeks after the outbreak we see a reaction from the number of answered contacts, it is more muted.

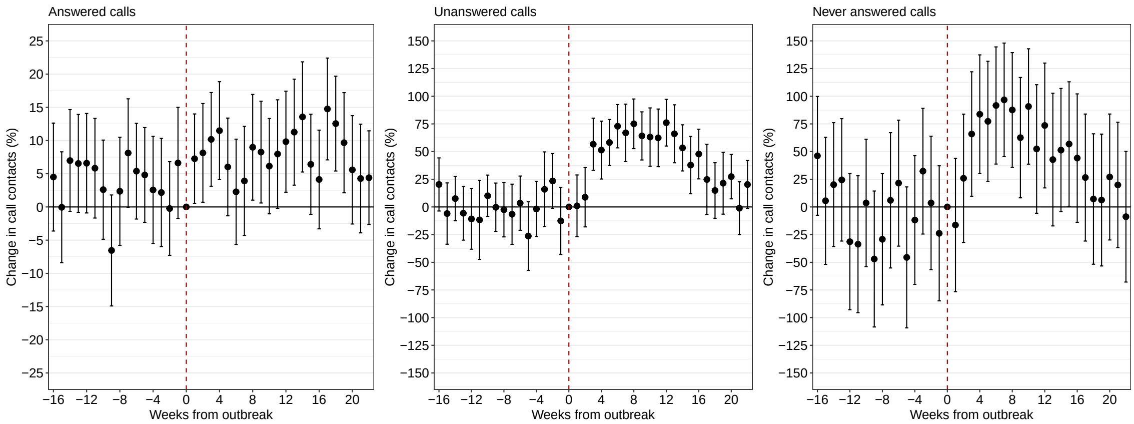


Figure 4.7: Event study results - capacity constraints during the first COVID-19 wave. Notes: The graph shows point estimates together with 95% confidence intervals from the weekly event study regression model (4.2) for the first COVID-19 wave for answered (left), unanswered (middle) and never answered (right) calls. Week zero serves as base week and lasts from 25 February 2020 to 02 March 2020. All coefficients are estimated relative to the base week. Note that the panel for answered calls (left) has a different scaling for better visibility. Robust standard errors are used for calculating uncertainty measures. Source: Network provider.

Our findings help explain the last section's result why total call duration does not expand in the same fashion as the total number of calls. Assuming that unanswered calls are underreported in the manual call listings from helpline call agents (which is in line with summary statistics), our result also gives an explanation why BL find no significant response after the outbreak of the pandemic in Switzerland.

4.5.4 Heterogeneity Analysis

Since helpline agents try to identify gender and age group of callers, we can analyze whether the pandemic has caused heterogeneous effects along those lines. In order to keep the sample size large, we refrain from analyzing callers identified as non-binary (less than 1000 calls in total). Given that the infection-fatality ratio is much higher among the elder population (Sorensen et al., 2022) and disaggregated age group samples are rather small, we run two separate regressions for callers above and below the age of 65 .[11]

Calling behavior conditional on gender

Table 4.4 presents evidence that male and female callers seek more mental support after the outbreak of the first COVID-19 wave. However, the number of contact attempts made by male callers is more severely affected by the pandemic. Depending on the underlying sample, men increase calling activity between 11 and 22% while women do so between 9 and 10%. This gender gap is also reflected in estimates for the total duration of calls, which increases between 18 and 29% for male callers, while being relatively muted for women (0-4%).

Therefore, our results contrast with the evidence presented in BL, who do not find a significant difference between men and women in Switzerland. Interestingly, most existing studies conclude that women's mental health is more heavily affected by the pandemic (Amerio et al., 2022, Davillas and Jones, 2021, Wang et al., 2021, Zamarro and Prados, 2021, Pieh et al., 2020, Rossi et al., 2020, see also the comprehensive literature review in Vloo et al., 2021). However, the main reason for this general finding may lie in the research design implemented in most studies. Most of the literature uses cross-sectional surveys during lockdowns or periods of high COVID-19 infection rates to find women suffering from a worse mental state. The fact that there exists a gender gap concerning depression, anxiety, and stress-related disorders even before the outbreak of the pandemic (Gao et al., 2020, Riecher-Rössler, 2017) should clearly be considered when interpreting recent cross-sectional results. Differences in mental health are likely to have existed before the pandemic,

[11] The available age categories are 0-18, 19-40, 41-65, and older than 65.

Dependent Variable:	ln(Calls)				ln(Duration)			
Coefficient								
$\hat{\beta}$	0.110***	0.102***	0.217***	0.089***	0.179***	0.042***	0.291***	0.001
	(0.027)	(0.012)	(0.056)	(0.023)	(0.035)	(0.011)	(0.076)	(0.022)
Gender								
Male callers	✓		✓		✓		✓	
Female callers		✓		✓		✓		✓
Analysis period								
01.01.2017 - 20.7.2020	✓	✓			✓	✓		
08.10.2019 - 20.7.2020			✓	✓			✓	✓
Pretreatment mean	25	49.8	27.3	53.2	162.1	390.5	179.9	416.4
Observations	15,564	15,564	3,444	3,444	15,564	15,564	3,444	3,444
R^2	0.51	0.71	0.60	0.76	0.55	0.73	0.60	0.83

Table 4.4. Pre-post study: calling behavior during the first COVID-19 wave conditional on gender. Notes: The table shows estimates from pre-post regression model (4.1) for the first COVID-19 wave for male and female callers. $\hat{\beta}$ is the coefficient from the indicator variable Post_t, which is set zero before 25 February 2020 and one afterwards. The model is estimated using data from a) 01 January 2017 to 20 July 2020 (i.e. ending 20 weeks after the reference day) as well as b) a symmetric time window of 20 weeks around the reference day. The first four rows show results for total call count as dependent variable, while subsequent rows do so for call duration. Robust standard errors are given in parentheses. Significance codes: ***: 0.01, **: 0.05, *: 0.1. Source: Network provider, DH.

thus calling for longitudinal research designs. A rare example is Vloo et al., 2021, which find women to be more affected by depression while men suffer more from anxiety. However, they compare dedicated COVID-19 questionnaires in the middle of the first outbreak with the last available survey wave, which is collected in the years 2014-2017.

Our finding that the pandemic tends to have put a relatively larger strain on men's mental health is thus a novel insight.[12] Strong help seeking behavior is even more remarkable given that men are known to underutilize counseling services (Mackenzie et al., 2006, Rith-Najarian et al., 2019).

[12]The gender gap in total call contacts is also present in our data since about two-thirds of all contacts come from female callers. For more descriptive statistics, see Table 4.11.

Calling behavior conditional on age group

Table 4.5 presents evidence on how the first COVID-19 wave affected mental health of people older and younger than 65 years. Since vulnerability to the virus increases with age, we expect people over 65 to seek relatively more help from call agents. In line with intuition, the increase in call counts is pronounced and lies between 24 and 29% for this group. The reaction of younger callers is relatively muted with 7% more contacts compared to before the pandemic. Turning to the effective talking duration, we find elderly callers to spend between 8 and 14% more time on the phone, although the first estimate is insignificant. Callers with an age lower than 65 increase the duration of the call between 4% and 9%, while the first estimate is again characterized by high statistical uncertainty.

Dependent Variable:	ln(Calls)				ln(Duration)			
Coefficient								
$\hat{\beta}$	0.069***	0.237***	0.066***	0.287***	0.092***	0.144***	0.044*	0.082
	(0.012)	(0.038)	(0.024)	(0.079)	(0.014)	(0.048)	(0.024)	(0.107)
Age								
Below 65	✓		✓		✓		✓	
65 and above		✓		✓		✓		✓
Analysis period								
01.01.2017 - 20.7.2020	✓	✓			✓	✓		
08.10.2019 - 20.7.2020			✓	✓			✓	✓
Pretreatment mean	55.6	15.7	61.4	15.3	420.2	112.7	463.3	111.1
Observations	15,564	15,564	3,444	3,444	15,564	15,564	3,444	3,444
R^2	0.69	0.58	0.78	0.53	0.69	0.62	0.80	0.60

Table 4.5. Pre-post study: calling behavior during the first COVID-19 wave conditional on age. Notes: The table shows estimates from pre-post regression model (4.1) for the first COVID-19 wave for callers identified as older or younger than 65 years. $\hat{\beta}$ is the coefficient from the indicator variable Post_t, which is set zero before 25 February 2020 and one afterwards. The model is estimated using data from a) 01 January 2017 to 20 July 2020 (i.e. ending 20 weeks after the reference day) as well as b) a symmetric time window of 20 weeks around the reference day. The first four rows show results for total call count as dependent variable, while subsequent rows do so for call duration. Robust standard errors are given in parentheses. Significance codes: ***: 0.01, **: 0.05, *: 0.1. Source: Network provider, DH.

The evidence that elderly people's mental health suffered stronger does not confirm the findings of the survey literature. Based on cross-sectional data sets, younger adults are found to be more severely affected by the pandemic than older individuals (Czeisler et al., 2020, Pieh et al., 2020, Bruine de Bruin, 2021, Wilson et al., 2021). However, our results are in line with the outcome of the longitudinal study in BL. For persons older than 65, they find an increase in call numbers of 29% during the first wave relative to the corresponding 2019 period.[13]. Similarly, Altindag et al. (2022) find a decrease in mental health for the elderly above 65 in Turkey after the introduction of social distancing laws for that age group.

4.6 Robustness

To shield against possible model misspecification we follow Silva and Tenreyro (2006) and estimate Poisson count models on the regional and also on the individual level. Since the latter involves a data set that would be characterized by a large number of zero observations for both call count and duration, we aggregate the data to a weekly frequency.[14] Finally, we show that estimating the baseline regression model (4.1) on an artificial first COVID-19 wave in 2019 does not replicate our results.

4.6.1 Weekly Effects

The regional model is specified as follows:

$$\ln\left[\mathbb{E}\left(\text{Calls}_{r,w}|\mathbf{X}\right)\right] = \beta\text{Post}_t + \mu_w + \eta_r \tag{4.4}$$

where w refers to the week of the call. The coefficient of interest β can be interpreted as the percentage change in the expected call contacts, allowing comparison with the results of baseline model (4.1) and DiD model (4.3).

[13] Interestingly, the pronounced response of elderly people during the first wave is not at all reflected in the second wave, as Appendix Table 4.10 shows.

[14] We have also estimated the regional model using daily observations. The results are similar to those of the baseline model (4.1) and the DiD model (4.3) and are available on request.

Table 4.6 shows that the estimation outcome is highly robust to model choice. The pandemic increases the expected number of call contacts between 6 and 10% (8-11% in the baseline and DiD models). The estimated coefficients for the effective duration of the call indicate an increase of approximately 3-4% after the outbreak (5% in the baseline and DiD models). Thus, our results extend to the Poisson count model.[15]

Dependent Variable:	Calls		Duration	
Coefficient				
$\hat{\beta}$	0.061***	0.098***	0.032***	0.038**
	(0.014)	(0.027)	(0.009)	(0.018)
Analysis period				
01 January 2017 - 20 July 2020	✓		✓	
08 October 2019 - 20 July 2020		✓		✓
Pretreatment mean	584.6	622.7	4124.5	4450.4
Observations	2,232	492	2,232	492
Pseudo R^2	0.77	0.83	0.91	0.94

Table 4.6. Pre-post study: Poisson count model for the first COVID-19 wave. Notes: The table shows estimates from pre-post Poisson regression model (4.4) for the first COVID-19 wave. $\hat{\beta}$ is the coefficient from the indicator variable Post_t, which is set zero before 25 February 2020 and one afterwards. The model is estimated using data from a) 01 January 2017 to 20 July 2020 (i.e. ending 20 weeks after the reference day) as well as b) a symmetric time window of 20 weeks around the reference day. The first two rows show results for call count as dependent variable, while subsequent rows do so for call duration. Robust standard errors are given in parentheses. Significance codes: ***: 0.01, **: 0.05, *: 0.1. Source: Network provider.

4.6.2 Individual Level Effects

So far, all models have been estimated on the regional level. Next we look at the individual level. The panel is constructed so that each number has weekly observations from the first to the last call, and if a caller is not calling in a given

[15]Estimating a negative binomial regression leads to similar results.

week the entry is equal to zero. Using the information on calling numbers, we estimate the following individual fixed-effects Poisson model:

$$\ln\left[\mathbb{E}\left(\text{Calls}_{i,w}|\mathbf{X}\right)\right] = \beta\text{Post}_t + \mu_w + \gamma_i \tag{4.5}$$

where the dependent variable is the number or duration of calls from phone number i during week w. We again include a linear time trend μ_w and add individual fixed-effects γ_i. In order to estimate fixed-effects for a large amount of numbers, we do so only for the entire sample spanning from January 2017 to 20 July 2020. For this period of time, the data set consists of calls from 69,212 unique phone numbers.[16] Table (4.7) shows that the coefficient estimates are larger than those of the aggregated regional model (4.4) estimated on the same sample. On average, people who have already called before the outbreak of the pandemic increased call contacts by 18.5%, while the increase in total duration is slightly muted at 15.7%.

4.6.3 Placebo Analysis for 2019

This section shows how the findings from the first COVID-19 wave in 2020 compare to a placebo wave one year earlier. We thus rerun regression model (4.1) with the only difference that the indicator variable Post_t is set equal to 1 for all days after 25 February 2019 (instead of 2020). Table (4.8) indeed shows that we do not find an increase of call contacts or duration when pretending 2019 to be the outbreak year of the pandemic.

4.7 Discussion and Conclusion

This article uses Swiss data from a call hotline that provides free counseling to identify the mental health effects of COVID-19 and associated social distancing measures. We estimate the effects on the number and duration of calls by combining

[16]From this set of numbers, roughly half (34,888) have called more than once and 13,438 have called before and after the pandemic. The latter are responsible for 75% of all calls between January 2017 and 20 July 2020.

Dependent Variable:	Calls	Duration
Coefficient		
$\hat{\beta}$	0.185***	0.157***
	(0.014)	(0.012)
Analysis period		
01 January 2017 - 20 July 2020	✓	✓
Pretreatment mean	0.6	3.9
Observations	2,351,527	2,277,610
Pseudo R^2	0.65	0.66

Table 4.7. Pre-post study: Individual Poisson count model for the first COVID-19 wave. Notes: The table shows estimates from pre-post individual fixed-effects Poisson regression model (4.5) for the first COVID-19 wave. $\hat{\beta}$ is the coefficient from the indicator variable Post_t, which is set zero before 25 February 2020 and one afterwards. The model is estimated using data from 01 January 2017 to 20 July 2020 (i.e. ending 20 weeks after the reference day). The first row shows the results for call count as dependent variable, while the second row does so for call duration. Robust standard errors are given in parentheses. Significance codes: ***: 0.01, **: 0.05, *: 0.1. Source: Network provider.

administrative data from the network provider with the call listings from the hotline. In particular, we analyze the two COVID-19 waves starting on 25 February 2020 and 19 October 2020, respectively, to estimate pre-post models and difference-in-differences models using 2019 as the control group.

We find significant increases in the number and duration of calls during the first wave. Moreover, we analyze unanswered calls and find that there was excessive demand for counseling that could not be met due to capacity constraints, thus increasing the unmet need for mental health. Our results also show that the strong increase in call activity is driven mainly (but not only) by people above the age of 65. In contrast to the cross-sectional literature, we find men seeking more help compared to female callers in Switzerland. On the other hand, we don't find significant effects for the second wave. These results are robust to an exponential transformation and including caller fixed effects. Finally, we introduce a placebo COVID-19 wave in 2019 and do not find any significant effects.

Dependent Variable:	ln(Calls)		ln(Duration)	
Coefficient				
$\hat{\beta}$	-0.017*	0.002	-0.022***	0.022
	(0.010)	(0.019)	(0.007)	(0.015)
Analysis period				
01 January 2017 - 21 July 2019	✓		✓	
08 October 2018 - 21 July 2019		✓		✓
Pretreatment mean	81	87.8	577.6	607.9
Observations	11,184	3,444	11,184	3,444
R^2	0.61	0.65	0.75	0.78

Table 4.8. Pre-post study: an artificial first COVID-19 wave in 2019. Notes: The table shows estimates from pre-post regression model (4.1) for an artificial first COVID-19 wave in 2019. $\hat{\beta}$ is the coefficient from the indicator variable Post_t, which is set zero before 25 February 2019 and one afterwards. Model (4.1) is estimated using data from a) 01 January 2017 to 21 July 2019 (i.e. ending 20 weeks after the reference day) as well as b) a symmetric time window of 20 weeks around the reference day. The first two rows show results for call count as dependent variable, while subsequent rows do so for call duration. Robust standard errors are given in parentheses. Significance codes: ***: 0.01, **: 0.05, *: 0.1. Source: Network provider.

Our findings address important gaps in the literature on the mental health effects of social distancing measures. Although effective in reducing infection rates, the mental health effects of such policies represent an important component that should be taken into account. Switzerland is a country that implemented less restrictive social distancing measures during the second wave, which could explain the absence of a mental health effect for the second wave. This is in contrast to neighboring countries such as France and Germany, which implemented stricter measures and where BKLR find significant effects for the second wave.

As governments around the world might return to social distancing measures, more empirical evidence on the effects on mental health is needed.

4.A Appendix

4.A.1 Figures

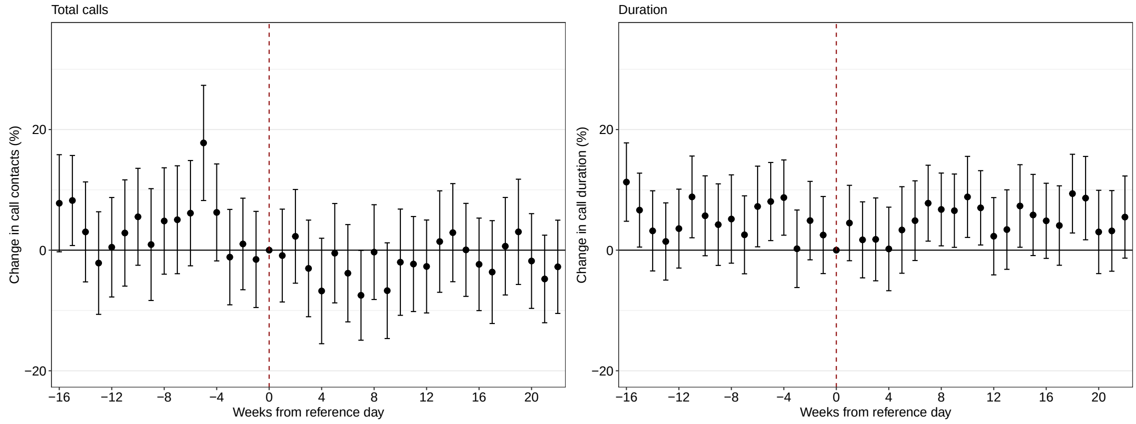


Figure 4.8: Event study: main findings for the second Covid-19 wave. Notes: The graph shows point estimates together with 95% confidence intervals from the weekly event study regression model (4.2) for the second Covid-19 wave. Week zero serves as base week and lasts from 19 October 2020 to 25 October 2020. All coefficients are estimated relative to the base week. Robust standard errors are used for calculating uncertainty measures. Source: Network provider.

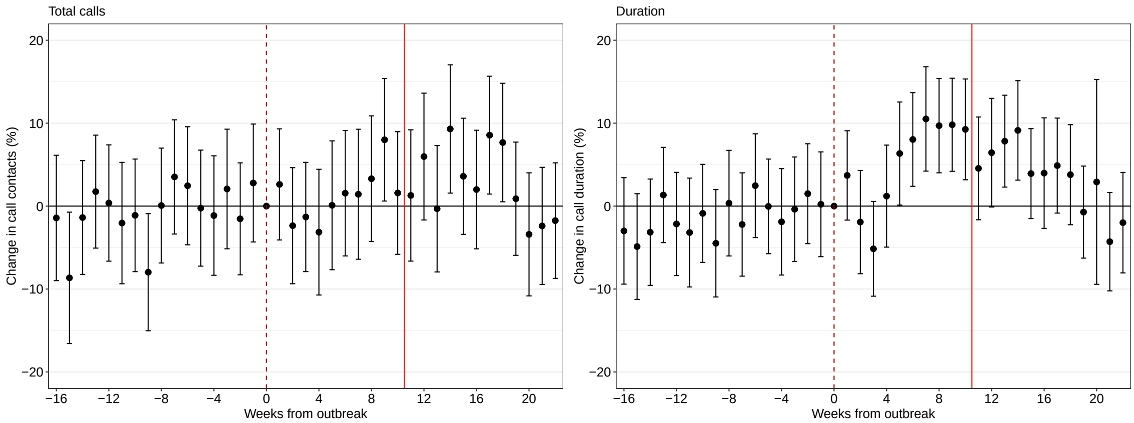


Figure 4.9: Event study: reproducing results in BL using helpline data. Notes: The graph shows point estimates together with 95% confidence intervals from the weekly event study regression model (4.2) for the first Covid-19 wave using helpline listings instead of network provider data. The solid red line after week 10 shows the end of the sample used in BL. Note that BL only analyse call volumes, but not duration. Week zero serves as base week and lasts from 25 February 2020 to 02 March 2020. All coefficients are estimated relative to the base week. Robust standard errors are used for calculating uncertainty measures. Source: DH.

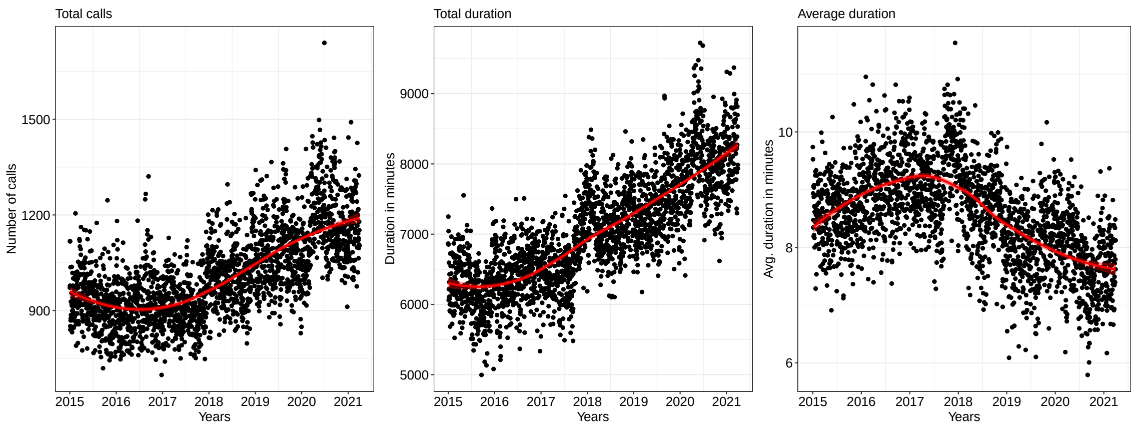


Figure 4.10: Nationwide call number and duration with trend. Notes: For the entire observation period, this figure shows scatter plots for daily call number (left), duration (middle) and average duration per call (right). Each plot is equipped with a red trend line which was computed using local polynomial regression. Source: Network provider.

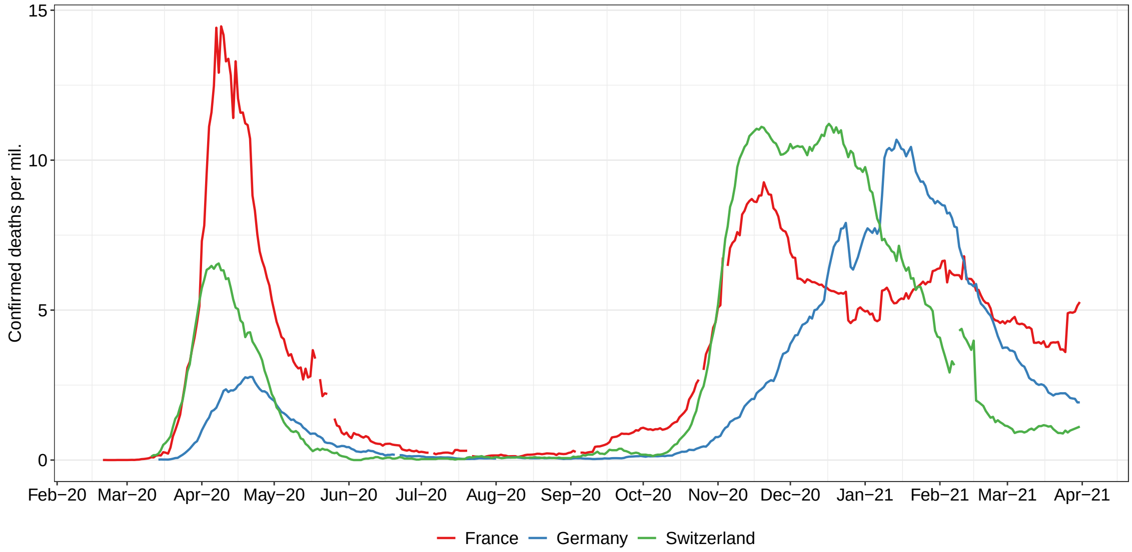


Figure 4.11: Covid-19 confirmed deaths. Notes: This figure shows daily confirmed deaths per million people for Switzerland, Germany and France. All values are plotted as seven day moving averages to reduce variability. Source: Ritchie et al. (2020).

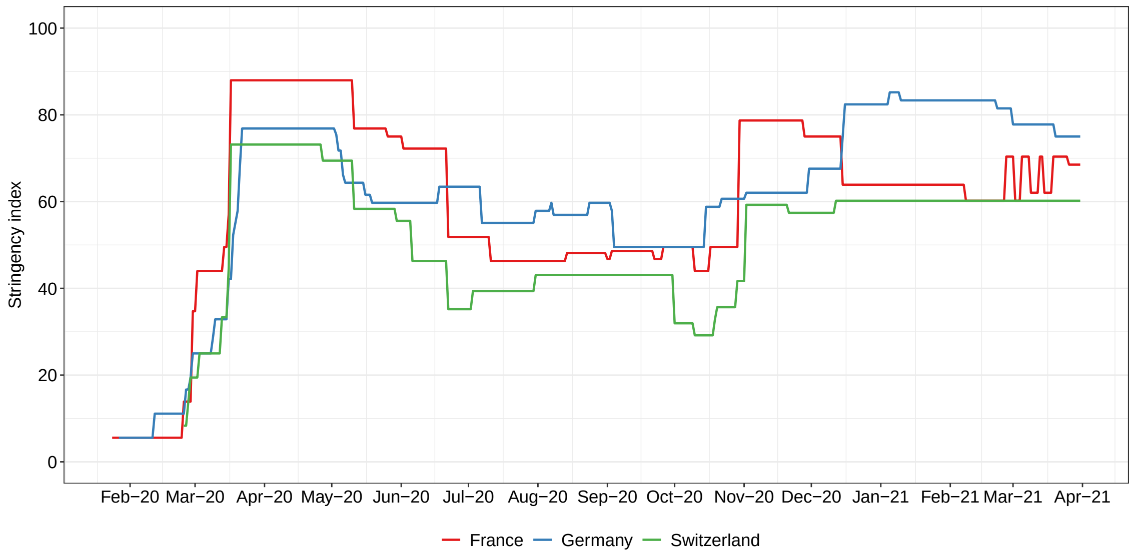


Figure 4.12: Stringency indices. Notes: This figure shows daily stringency indices for Switzerland, Germany and France from the Oxford Covid-19 government response tracker. The stringency index ranges from 0 to 100 with the latter indicating the most restrictive policy regime. Source: Ritchie et al. (2020).

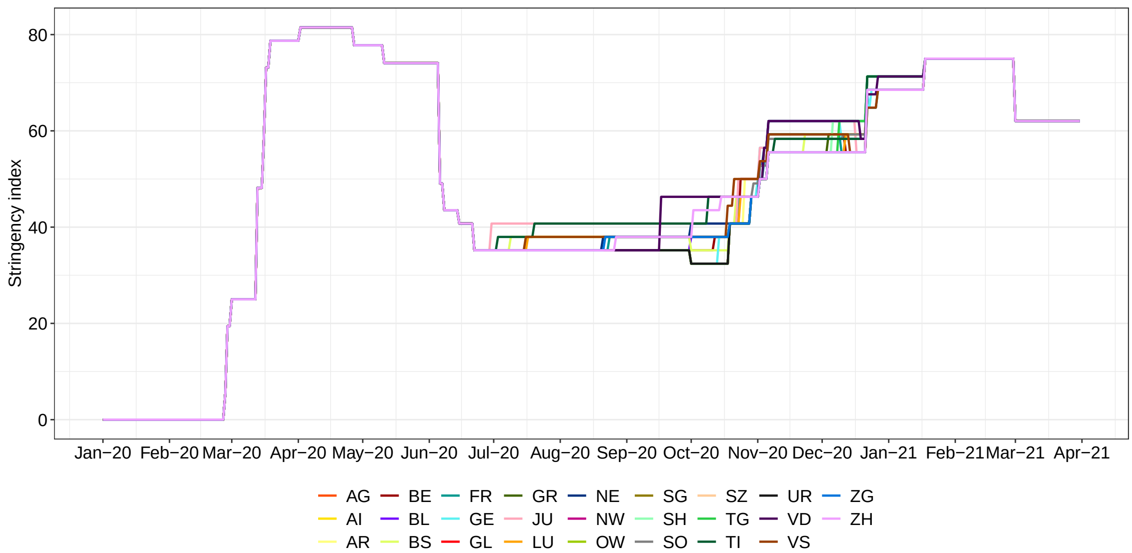


Figure 4.13: Cantonal stringency indices. Notes: This figure shows daily stringency indices for each canton in Switzerland. The stringency index ranges from 0 to 100 with the latter indicating the most restrictive policy regime. Source: Pleninger et al. (2022).

4.A.2 Tables

Dependent Variable:	ln(Calls)		ln(Duration)	
Coefficient				
$\hat{\beta}$	-0.011	-0.003	0.036***	0.041***
	(0.008)	(0.016)	(0.009)	(0.015)
Analysis period				
01 January 2017 - 20 July 2020	✓		✓	
08 October 2019 - 20 July 2020		✓		✓
Pretreatment mean	57	59.3	643.5	687
Observations	11,593	2,583	11,593	2,583
R^2	0.76	0.78	0.61	0.80

Table 4.9. Pre-post study: reproducing results in BL using helpline data. Notes: The table shows estimates from pre-post regression model (4.1) for the first Covid-19 wave using helpline listings instead of network provider data. This allows to reproduce the results found in BL. Note that BL only analyse call volumes, but not duration. $\hat{\beta}$ is the coefficient from the indicator variable Post_t, which is set zero before 25 February 2020 and one afterwards. The model is estimated using data from a) 01 January 2017 to 20 July 2020 (i.e. ending 20 weeks after the reference day) as well as b) a symmetric time window of 20 weeks around the reference day. The first two rows show results for call count as dependent variable, while subsequent rows do so for call duration. Robust standard errors are given in parentheses. Significance codes: ***: 0.01, **: 0.05, *: 0.1. Source: DH.

Dependent Variable:	ln(Calls)		ln(Duration)	
Coefficient				
$\hat{\beta}$	-0.068*	0.020	0.045	-0.008
	(0.038)	(0.062)	(0.048)	(0.083)
Age				
Below 65	✓		✓	
65 and above		✓		✓
Analysis period				
01 June 2020 - 14 March 2021	✓	✓	✓	✓
Pretreatment mean	71	16.3	502.8	104.4
Observations	3,444	3,444	3,444	3,444
R^2	0.71	0.47	0.69	0.52

Table 4.10. Pre-post study: calling behavior during the second Covid-19 wave conditional on age. Notes: The table shows estimates from pre-post regression model (4.1) for the second Covid-19 wave for callers identified as older or younger than 65 years. $\hat{\beta}$ is the coefficient from the indicator variable Post_t, which is set zero before 19 October 2020 and one afterwards. The model is estimated using data from 01 June 2020 to 14 March 2021 (i.e. ending 20 weeks after the reference day). The first two rows show results for total call count as dependent variable, while subsequent rows do so for call duration. Robust standard errors are given in parentheses. Significance codes: ***: 0.01, **: 0.05, *: 0.1. Source: Network provider, DH.

	Male	Female	Diverse	Unclear	NA	**Sum**
-18	9,203	8,062	2	78	0	17,345
19-40	118,293	149,315	0	13	0	267,621
41-65	298,082	513,684	4	32	0	811,802
65+	35,591	261,850	0	6	0	297,447
Unclear	30,062	40,247	4	797	0	71,110
NA	0	0	0	0	167,600	167,600
Sum	491,231	973,158	10	926	167,600	1,632,925

Table 4.11. Summary table age and gender. Notes: The table shows the number of observations in each gender/age group. Source: Network provider, DH.

	2017	2018	2019	2020	2021	**Sum**
AG	15,627	25,939	21,664	26,797	6,586	96,613
BE	35,828	36,377	41,895	47,595	11,137	172,832
BI	25,337	30,452	29,555	29,563	6,744	121,651
BS	16,828	17,778	18,269	21,422	5,193	79,490
GE	24,955	24,383	24,453	27,325	6,324	107,440
LU	18,577	20,072	24,339	24,603	6,492	94,083
SG	25,271	29,602	28,096	33,004	8,772	124,745
TI	30,501	30,605	28,919	29,214	8,044	127,283
VD	42,464	45,745	39,970	48,047	11,085	187,311
VS	27,934	25,321	30,683	32,888	7,281	124,107
WI	17,674	20,051	25,560	28,226	7,312	98,823
ZH	53,057	64,633	79,515	81,175	20,073	298,453
NA	9	17	20	36	12	94
Sum	334,062	370,975	392,938	429,895	105,055	1,632,925

Table 4.12. Summary table regional centers and years. Notes: The table shows the number of observations for each regional helpline center per year for the network provider data set. Regional abbreviations are as follows: AG = Aargau, BE = Bern, BI = Biel/Bienne, BS = Basel-Stadt, GE = Geneve, LU = Lucerne, SG = St. Gallen, TI = Ticino, VD = Vaud, VS = Valais, WI = Winterthur, ZH = Zurich. Source: Network provider.

		Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Helpline							
	All	0.0	0.0	7.9	11.5	18.8	22,220.0
	At least 10 sec.	0.2	6.1	14.0	16.2	22.8	22,220.0
Network provider							
	All	0.0	0.1	0.6	7.0	10.5	720.9
	At least 10 sec.	0.2	0.6	5.3	10.4	16.6	720.9

Table 4.13. Summary table call duration. Notes: The table shows summary statistics for call duration in minutes. All figures are computed using either all calls (first row) or only calls which exceed a duration of ten seconds (second row). Source: Network provider, DH.

Bibliography

Adrian, T., E. Moench, and H. Shin (2010). Financial intermediation, asset prices and macroeconomic dynamics. *FRB of New York Staff Report* (422).

Ahn, S. and A. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*(3), 1203–1227.

Aladangady, A. (2017). Housing wealth and consumption: Evidence from geographically-linked microdata. *American Economic Review 107*(11), 3415–46.

Alessi, L., M. Barigozzi, and M. Capasso (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics and Probability Letters 80*(23-24), 1806–1813.

Altindag, O., B. Erten, and P. Keskin (2022). Mental health costs of lockdowns: Evidence from age-specific curfews in Turkey. *American Economic Journal: Applied Economics 14*(2), 320–343.

Amerio, A., P. Bertuccio, F. Santi, D. Bianchi, A. Brambilla, A. Morganti, A. Odone, A. Costanza, C. Signorelli, A. Aguglia, et al. (2022). Gender differences in COVID-19 lockdown impact on mental health of undergraduate students. *Frontiers in Psychiatry*, 2421.

Amir-Ahmadi, P. and H. Uhlig (2015). Sign restrictions in Bayesian FAVARs with an application to monetary policy shocks. *NBER Working Paper* (21738).

Anderes, M., A. Rathke, S. Streicher, and J.-E. Sturm (2021). The role of ECB communication in guiding markets. *Public Choice 186*(3-4), 351–383.

Andersen, H. and S. Leth-Petersen (2019). Housing wealth effects and mortgage borrowing: The effect of subjective unanticipated changes in home values on home equity extraction in Denmark. *CEPR Discussion Paper* (13926).

André, C., R. Gupta, and P. Kanda (2012). Do house prices impact consumption and interest rate? *OECD Working Papers* (947).

Angrist, J. D. and J.-S. Pischke (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives 24*(2), 3–30.

Armbruster, S. and V. Klotzbücher (2020). Lost in lockdown? COVID-19, social distancing, and mental health in Germany. *University of Freiburg Discussion Paper Series* (2020-04).

Aruoba, B., F. Diebold, J. Nalewaik, F. Schorfheide, and D. Song (2016). Improving GDP measurement: A measurement-error perspective. *Journal of Econometrics 191*(2), 384–397.

Ayouni, I., J. Maatoug, W. Dhouib, N. Zammit, S. Fredj, R. Ghammam, and H. Ghannem (2021). Effective public health measures to mitigate the spread of COVID-19: A systematic review. *BMC Public Health 21*(1), 1–14.

Bagliano, F. and C. Morana (2012). The Great Recession: US dynamics and spillovers to the world economy. *Journal of Banking and Finance 36*(1), 1–13.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bai, J. and S. Ng (2008a). Forecasting economic time series using targeted predictors. *Journal of Econometrics 146*(2), 304–317.

Bai, J. and S. Ng (2008b). Large dimensional factor analysis. *Foundations and Trends in Econometrics 3*(2), 89–163.

Bai, J. and P. Wang (2015). Identification and Bayesian estimation of dynamic factor models. *Journal of Business and Economic Statistics 33*(2), 221–240.

Baird, S., J. De Hoop, and B. Özler (2013). Income shocks and adolescent mental health. *Journal of Human Resources 48*(2), 370–403.

Banks, J. and X. Xu (2020). The mental health effects of the first two months of lockdown during the COVID-19 pandemic in the UK. *Fiscal Studies 41*(3), 685–708.

Barbarino, A., T. Berge, H. Chen, and A. Stella (2020). Which output gap estimates are stable in real time and why? *FRB Finance and Economics Discussion Series* (10).

Baron, J., E. Goldstein, and C. Wallace (2020). Suffering in silence: How COVID-19 school closures inhibit the reporting of child maltreatment. *Journal of Public Economics 190*, 104258.

Batchelor, S., S. Stoyanov, J. Pirkis, and K. Kõlves (2021). Use of kids helpline by children and young people in Australia during the COVID-19 pandemic. *Journal of Adolescent Health 68*(6), 1067–1074.

Batty, M., J. Bricker, J. Briggs, A. Volz, E. Holmquist, S. McIntosh, K. Moore, E. Nielsen, S. Reber, M. Shatto, and K. Sommer (2019). Introducing the distributional financial accounts of the United States. *Finance and Economics Discussion Series* (17).

Baxter, M. and R. King (1999). Measuring business cycles: Approximate band-pass filters for economic time series. *Review of Economics and Statistics 81*(4), 575–593.

Belviso, F. and F. Milani (2006). Structural factor-augmented VARs (SFAVARs) and the effects of monetary policy. *The BE Journal of Macroeconomics 6*(3).

Berger, D., V. Guerrieri, G. Lorenzoni, and J. Vavra (2017). House prices and consumer spending. *The Review of Economic Studies 85*(3), 1502–1542.

Berger, H., J. De Haan, and J.-E. Sturm (2011). Does money matter in the ECB strategy? New evidence based on ECB communication. *International Journal of Finance and Economics 16*(1), 16–31.

Bernanke, B., J. Boivin, and P. Eliasz (2005). Measuring the effects of monetary policy: A factor augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics 120*(1), 387–422.

Bernholz, P. (2015). Monetary regimes and inflation: History, economic and political relationships. In *Monetary Regimes and Inflation*. Edward Elgar Publishing.

Bernoth, K. and J. v. Hagen (2004). The Euribor futures market: Efficiency and the impact of ECB policy announcements. *International Finance 7*(1), 1–24.

Black, F. (1995). Interest rates as options. *The Journal of Finance 50*(5), 1371–1376.

Blanchard, O. and D. Quah (1989). The dynamic effects of aggregate demand and supply disturbances. *American Economic Review 79*(4), 655–673.

Blinder, A., M. Ehrmann, J. De Haan, and D.-J. Jansen (2017). Necessity as the mother of invention: Monetary policy after the crisis. *Economic Policy 32*(92), 707–755.

Blinder, A., M. Ehrmann, M. Fratzscher, J. De Haan, and D.-J. Jansen (2008). Central bank communication and monetary policy: A survey of theory and evidence. *Journal of Economic Literature 46*(4), 910–945.

Boivin, J., M. Giannoni, and D. Stevanović (2018). Dynamic effects of credit shocks in a data-rich environment. *Journal of Business and Economic Statistics*, 1–13.

Boivin, J. and S. Ng (2006). Are more data always better for factor analysis? *Journal of Econometrics 132*(1), 169–194.

Borio, C. and L. Gambacorta (2017). Monetary policy and bank lending in a low interest rate environment: diminishing effectiveness? *Journal of Macroeconomics 54*, 217–231.

Bostic, R., S. Gabriel, and G. Painter (2009). Housing wealth, financial wealth, and consumption: New evidence from micro data. *Regional Science and Urban Economics 39*(1), 79–89.

Bound, J., C. Brown, and N. Mathiowetz (2001). Measurement error in survey data. In *Handbook of Econometrics*, Volume 5, pp. 3705–3843. Elsevier.

Brand, C., D. Buncic, and J. Turunen (2010). The impact of ECB monetary policy decisions and communication on the yield curve. *Journal of the European Economic Association 8*(6), 1266–1298.

Bredin, D., S. Hyde, and G. Reilly (2010). Monetary policy surprises and international bond markets. *Journal of International Money and Finance 29*(6), 988–1002.

Brodeur, A., A. Clark, S. Fleche, and N. Powdthavee (2021). COVID-19, lockdowns and well-being: Evidence from Google trends. *Journal of Public Economics 193*, 104346.

Bruine de Bruin, W. (2021). Age differences in COVID-19 risk perceptions and mental health: Evidence from a national US survey conducted in March 2020. *The Journals of Gerontology: Series B 76*(2), e24–e29.

Brülhart, M., V. Klotzbücher, R. Lalive, and S. Reich (2021). Mental health concerns during the COVID-19 pandemic as revealed by helpline calls. *Nature 600*(7887), 121–126.

Brülhart, M. and R. Lalive (2020). Daily suffering: Helpline calls during the COVID-19 crisis. *Covid Economics 19*, 143–158.

Buch, C., S. Eickmeier, and E. Prieto (2014). Macroeconomic factors and microlevel bank behavior. *Journal of Money, Credit and Banking 46*(4), 715–751.

Bulíř, A., M. Čihák, and D.-J. Jansen (2013). What drives clarity of central bank communication about inflation? *Open Economies Review 24*, 125–145.

Bullinger, L., J. Carr, and A. Packham (2021). COVID-19 and crime: Effects of stay-at-home orders on domestic violence. *American Journal of Health Economics 7*(3), 249–280.

Burdett, A., A. Davillas, and B. Etheridge (2021). Weather, mental health, and mobility during the first wave of the COVID-19 pandemic. *Health Economics 30*(9), 2296–2306.

Canova, F. and G. De Nicolo (2002). Monetary disturbances matter for business fluctuations in the G-7. *Journal of Monetary Economics 49*(6), 1131–1159.

Canova, F. and M. Paustian (2011). Business cycle measurement with some theory. *Journal of Monetary Economics 58*(4), 345–361.

Cardarelli, R., T. Monacelli, A. Rebucci, and L. Sala (2009). Housing finance, housing shocks and the business cycle: Evidence from OECD countries. *Unpublished manuscript*.

Carroll, C., M. Otsuka, and J. Slacalek (2011). How large are housing and financial wealth effects? A new approach. *Journal of Money, Credit and Banking 43*(1), 55–79.

Carter, C. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika 81*(3), 541–553.

Cesa-Bianchi, A. (2013). Housing cycles and macroeconomic fluctuations: A global perspective. *Journal of International Money and Finance 37*, 215–238.

Chan, J. and R. Strachan (2023). Bayesian state space models in macroeconometrics. *Journal of Economic Surveys 37*(1), 58–75.

Christiano, L., R. Motto, and M. Rostagno (2014). Risk shocks. *American Economic Review 104*(1), 27–65.

Cochrane, J., J. Taylor, and V. Wieland (2019). Evaluating rules in the Fed's report and measuring discretion. In *Conference, Strategies for Monetary Policy, Hoover Institution, Stanford University, May*, Volume 3.

Coenen, G., M. Ehrmann, G. Gaballo, P. Hoffmann, A. Nakov, S. Nardelli, E. Persson, and G. Strasser (2017). Communication of monetary policy in unconventional times. *ECB Working Paper* (2080).

Coibion, O. and Y. Gorodnichenko (2015). Is the Phillips curve alive and well after all? Inflation expectations and the missing disinflation. *American Economic Journal: Macroeconomics 7*(1), 197–232.

Coibion, O., Y. Gorodnichenko, L. Kueng, and J. Silvia (2017). Innocent bystanders? Monetary policy and inequality. *Journal of Monetary Economics 88*, 70–89.

Coibion, O., Y. Gorodnichenko, and M. Ulate (2018). The cyclical sensitivity in estimates of potential output. *Brookings Papers on Economic Activity* (Fall).

Conrad, C. and M. Lamla (2010). The high-frequency response of the EUR-USD exchange rate to ECB communication. *Journal of Money, Credit and Banking 42*(7), 1391–1417.

Constâncio, V. (2018, May). Past and future of the ECB monetary policy. In *Speech at Conference on Central Banks in Historical Perspective: What Changed after the Financial Crisis, Central Bank of Malta, Valletta*, Volume 4.

Costa-Font, J., M. Knapp, and C. Vilaplana-Prieto (2022). The 'welcomed lockdown' hypothesis? Mental wellbeing and mobility restrictions. *The European Journal of Health Economics*, 1–21.

Czeisler, M., R. Lane, E. Petrosky, J. Wiley, A. Christensen, R. Njai, M. Weaver, R. Robbins, E. Facer-Childs, L. Barger, et al. (2020). Mental health, substance use, and suicidal ideation during the COVID-19 pandemic. *Morbidity and Mortality Weekly Report 69*(32), 1049.

Davillas, A. and A. Jones (2021). The first wave of the COVID-19 pandemic and its impact on socioeconomic inequality in psychological distress in the UK. *Health Economics 30*(7), 1668–1683.

De Haan, J. (2008). The effect of ECB communication on interest rates: An assessment. *The Review of International Organizations 3*, 375–398.

De Haan, J. and J.-E. Sturm (2019). Central bank communication – how to manage expectations? In D. Mayes, P. Siklos, and J.-E. Sturm (Eds.), *The Oxford handbook of the economics of central banking*, Chapter 8, pp. 231–262. Oxford University Press.

Deb, P., D. Furceri, J. Ostry, and N. Tawk (2022). The economic effects of COVID-19 containment measures. *Open Economies Review 33*(1), 1–32.

Diebold, F. and R. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics 13*(3), 253–263.

Dotsey, M., S. Fujita, and T. Stark (2018). Do Phillips curves conditionally help to forecast inflation? *International Journal of Central Banking 14*(4), 43–92.

Drehmann, M., C. Borio, L. Gambacorta, G. Jiménez, and C. Trucharte (2010). Countercyclical capital buffers: Exploring options. *BIS Working Papers* (317), 1–58.

Edge, R. and J. Rudd (2016). Real-time properties of the Federal Reserve's output gap. *Review of Economics and Statistics 98*(4), 785–791.

Ehrmann, M. and M. Fratzscher (2007). Communication by central bank committee members: Different strategies, same effectiveness? *Journal of Money, Credit and Banking 39*(2-3), 509–541.

Ehrmann, M. and M. Fratzscher (2013). Dispersed communication by central bank committees and the predictability of monetary policy decisions. *Public Choice 157*, 223–244.

Erten, B., P. Keskin, and S. Prina (2022). Social distancing, stimulus payments, and domestic violence: Evidence from the US during COVID-19. *AEA Papers and Proceedings 112*, 262–266.

Fausch, J. and M. Sigonius (2018). The impact of ECB monetary policy surprises on the German stock market. *Journal of Macroeconomics 55*, 46–63.

Fernald, J. (2014). Productivity and potential output before, during, and after the Great Recession. *NBER Macroeconomics Annual 29*.

Fetzer, T., L. Hensel, J. Hermle, and C. Roth (2021). Coronavirus perceptions and economic anxiety. *Review of Economics and Statistics 103*(5), 968–978.

Filardo, A. and B. Hofmann (2014). Forward guidance at the zero lower bound. *BIS Quarterly Review*.

Fleischman, C. and J. Roberts (2011). From many series, one cycle: Improved estimates of the business cycle from a multivariate unobserved components model. *FRB Finance and Economics Discussion Series* (10).

Friedman, M. (1970, September). The counter-revolution in monetary theory: First Wincott memorial lecture. In *Speech at the Senate House, University of London*, Volume 33.

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

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis 15*(2), 183–202.

Furlanetto, F., K. Hagelund, F. Hansen, and Ø. Robstad (2023, 2). Norges Bank output gap estimates: Forecasting properties, reliability, cyclical sensitivity and hysteresis. *Oxford Bulletin of Economics and Statistics 85*, 238–267.

Furlanetto, F., F. Ravazzolo, and S. Sarferaz (2019). Identification of financial factors in economic fluctuations. *The Economic Journal 129*(617), 311–337.

Galardo, M. and C. Guerrieri (2017). The effects of central bank's verbal guidance: Evidence from the ECB. *Bank of Italy Temi di Discussione (Working Paper)* (1129).

Gao, W., S. Ping, and X. Liu (2020). Gender differences in depression, anxiety, and stress among college students: A longitudinal study from China. *Journal of Affective Disorders 263*, 292–300.

Garratt, A., J. Mitchell, and S. Vahey (2014). Measuring output gap nowcast uncertainty. *International Journal of Forecasting 30*(2), 268–279.

Gerlach, S. and F. Smets (1997). Output gaps and inflation: Unobservable-components estimates for the G-7 countries. mimeograph, Bank for International Settlements.

Gerlach, S. and F. Smets (1999). Output gaps and monetary policy in the EMU area. *European Economic Review 43*(4-6), 801–812.

Geweke, J. (1991). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Federal Reserve Bank of Minneapolis Staff Report 148*.

Gibson, G., M. Grignon, J. Hurley, and L. Wang (2019). Here comes the SUN: Self-assessed unmet need, worsening health outcomes, and health care inequity. *Health Economics 28*(6), 727–735.

Grant, A. and J. Chan (2017). Reconciling output gaps: Unobserved components model and Hodrick-Prescott filter. *Journal of Economic Dynamics and Control 75*, 114–121.

Grothe, M. and A. Meyler (2015). Inflation forecasts: Are marked-based and survey-based measures informative? *ECB Working Paper* (1865).

Guérin, P., L. Maurin, and M. Mohr (2015). Trend-cycle decomposition of output in Euro Area inflation forecasts: A real-time approach based on model combination. *Macroeconomic Dynamics 19*(2), 363–393.

Guren, A., A. McKay, E. Nakamura, and J. Steinsson (2020). What do we learn from cross-regional empirical estimates in macroeconomics? *NBER Working Paper* (26881).

Guren, A., A. McKay, E. Nakamura, and J. Steinsson (2021). Housing wealth effects: The long view. *The Review of Economic Studies 88*(2), 669–707.

Halford, E., A. Lake, and M. Gould (2020). Google searches for suicide and suicide risk factors in the early stages of the COVID-19 pandemic. *PloS one 15*(7), e0236777.

Hamilton, J. (2018). Why you should never use the HP filter. *The Review of Economic and Statistic 100*(5), 831–843.

Hartmann, P. and F. Smets (2018). The first twenty years of the European Central Bank: Monetary policy. *ECB Working Paper* (2219).

Harvey, D., S. Leybourne, and P. Newbold (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting 13*(2), 281–291.

Hodrick, R. and E. Prescott (1997). Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking 29*(1), 1–16.

Hollander, M., D. Wolfe, and E. Chicken (2015). *Nonparametric statistical methods.* John Wiley & Sons.

Iacoviello, M. (2005). House prices, borrowing constraints, and monetary policy in the business cycle. *American Economic Review 95*(3), 739–764.

Iacoviello, M. and S. Neri (2010). Housing market spillovers: Evidence from an estimated DSGE model. *American Economic Journal: Macroeconomics 2*(2), 125–164.

Inui, M., N. Sudou, and T. Yamada (2017). Effects of monetary policy shocks on inequality in Japan. *Bank of Japan Working Paper* (17-E-3).

Jacobs, J. P., S. Sarferaz, J. E. Sturm, and S. van Norden (2022). Can GDP measurement be further improved? Data revision and reconciliation. *Journal of Business and Economic Statistics 40*, 423–431.

Jarocinski, M. and F. Smets (2008). House prices and the stance of monetary policy. *ECB Working Paper* (891).

Jones, C. (2014). Bundesbank president Jens Weidmann steps up criticism of QE. *Financial Times*.

Justiniano, A., G. Primiceri, and A. Tambalotti (2010). Investment shocks and business cycles. *Journal of Monetary Economics 57*(2), 132–145.

Justiniano, A., G. Primiceri, and A. Tambalotti (2015). Household leveraging and deleveraging. *Review of Economic Dynamics 18*(1), 3–20.

Kamber, G., J. Morley, and B. Wong (2018). Intuitive and reliable estimates of the output gap from a Beveridge-Nelson filter. *Review of Economics and Statistics 100*(3), 550–566.

Kaplan, G., G. Violante, and J. Weidner (2014). The wealthy hand-to-mouth. *NBER Working Paper* (20073).

Kim, C.-J. and C. Nelson (1999). *State-space models with regime switching: Classical and Gibbs-sampling approaches with applications.* The MIT Press Series. MIT Press.

Kourentzes, N. and G. Athanasopoulos (2019). Cross-temporal coherent forecasts for Australian tourism. *Annals of Tourism Research 75*, 393–409.

Krippner, L. (2013a). Measuring the stance of monetary policy in zero lower bound environments. *Economics Letters 118*(1), 135–138.

Krippner, L. (2013b). A tractable framework for zero-lower-bound Gaussian term structure models. *CAMA Working Paper* (49).

Krippner, L. (2015). *Zero lower bound term structure modeling: A practitioner's guide.* Springer.

Kuttner, K. (1992). Monetary policy with uncertain estimates of potential output. *Economic Perspectives 16*(1), 2–15.

Kuttner, K. (1994). Estimating potential output as a latent variable. *Journal of Business and Economic Statistics 12*(3), 361–368.

Lamla, M. and J.-E. Sturm (2013). Interest rate expectations in the media and central bank communication. In P. Siklos and J.-E. Sturm (Eds.), *Central Bank Communication, Decision Making, and Governance*, Chapter 5, pp. 101–111. MIT Press.

Leamer, E. (1983). Let's take the con out of econometrics. *The American Economic Review 73*(1), 31–43.

Leamer, E. (2007). Housing is the business cycle. Technical report, National Bureau of Economic Research.

Leslie, E. and R. Wilson (2020). Sheltering in place and domestic violence: Evidence from calls for service during COVID-19. *Journal of Public Economics 189*, 104241.

Liu, Z., P. Wang, and T. Zha (2013). Land-price dynamics and macroeconomic fluctuations. *Econometrica 81*(3), 1147–1184.

Mackenzie, C., W. Gekoski, and J. Knox (2006). Age, gender, and the underutilization of mental health services: The influence of help-seeking attitudes. *Aging and Mental Health 10*(6), 574–582.

McCoy, E. and U. Clemens (2017). A calibration of the shadow rate to the Euro area using genetic algorithms. *ECB Discussion Paper* (51).

McCracken, M. and S. Ng (2020). Fred-QD: A quarterly database for macroeconomic research. Technical report, National Bureau of Economic Research.

McInerney, M., J. Mellor, and L. Nicholas (2013). Recession depression: Mental health effects of the 2008 stock market crash. *Journal of Health Economics 32*(6), 1090–1104.

Mendez-Lopez, A., D. Stuckler, M. McKee, J. Semenza, and J. Lazarus (2022). The mental health crisis during the COVID-19 pandemic in older adults and the role of physical distancing interventions and social protection measures in 26 European countries. *SSM-Population Health 17*, 101017.

Mian, A., K. Rao, and A. Sufi (2013). Household balance sheets, consumption, and the economic slump. *The Quarterly Journal of Economics 128*(4), 1687–1726.

Mian, A. and A. Sufi (2011). House prices, home equity-based borrowing, and the US household leverage crisis. *American Economic Review 101*(5), 2132–56.

Mian, A. and A. Sufi (2014). What explains the 2007-2009 drop in employment? *Econometrica 82*(6), 2197–2223.

Mise, E., T.-H. Kim, and P. Newbold (2005). On suboptimality of the Hodrick–Prescott filter at time series endpoints. *Journal of Macroeconomics 27*(1), 53–67.

Monreal-Bartolomé, A., Y. López-Del-Hoyo, I. Cabrera-Gil, A. Aguilar-Latorre, M. Puebla-Guedea, S. Boira, and J. Lanero (2022). Analysis of the calls received during the COVID-19 lockdown by the mental health crisis helpline operated by the Professional College of Psychology of Aragon. *International Journal of Environmental Research and Public Health 19*(5), 2901.

Morley, J. and J. Piger (2012). The asymmetric business cycle. *Review of Economics and Statistics 94*(1), 208–221.

Muellbauer, J. (2008). Housing, credit and consumer expenditure. *CEPR Discussion Paper* (6782).

Mumtaz, H. and A. Theophilopoulou (2017). The impact of monetary policy on inequality in the UK: An empirical analysis. *European Economic Review 98*, 410–423.

Muresan, G.-M., V.-L. Văidean, C. Mare, and M. Achim (2022). Were we happy and we didn't know it? A subjective dynamic and financial assessment pre-, during and post-COVID-19. *The European Journal of Health Economics*, 1–20.

Musso, A., S. Neri, and L. Stracca (2011). Housing, consumption and monetary policy: How different are the US and the Euro area? *Journal of Banking and Finance 35*(11), 3019–3041.

Neuenkirch, M. (2013). Monetary policy transmission in vector autoregressions: A new approach using central bank communication. *Journal of Banking and Finance 37*(11), 4278–4285.

Orphanides, A. (2003a). Monetary policy evaluation with noisy information. *Journal of Monetary Economics 50*(3), 605–631.

Orphanides, A. (2003b). The quest for prosperity without inflation. *Journal of Monetary Economics 50*(3), 633–663.

Orphanides, A. and S. van Norden (2002). The unreliability of output-gap estimates in real time. *The Review of Economics and Statistics 84*(4), 569–583.

Orphanides, A. and S. van Norden (2005). The reliability of inflation forecasts based on output gap estimates in real time. *Journal of Money, Credit and Banking 37*(3), 583–601.

Pettenuzzo, D. and A. Timmermann (2017). Forecasting macroeconomic variables under model instability. *Journal of Business and Economic Statistics 35*(2), 183–201.

Phillips, J. and C. Nugent (2014). Suicide and the Great Recession of 2007–2009: The role of economic factors in the 50 US states. *Social Science and Medicine 116*, 22–31.

Pichette, L., M. Robitaille, M. Salameh, and P. St-Amant (2019). Dismiss the output gaps? To use with caution given their limitations. *Economic Modelling 76*, 199–215.

Pieh, C., S. Budimir, and T. Probst (2020). The effect of age, gender, income, work, and physical activity on mental health during coronavirus disease (COVID-19) lockdown in Austria. *Journal of Psychosomatic Research 136*, 110186.

Pirkis, J., A. John, S. Shin, M. DelPozo-Banos, V. Arya, P. Analuisa-Aguilar, L. Appleby, E. Arensman, J. Bantjes, A. Baran, et al. (2021). Suicide trends in the early months of the COVID-19 pandemic: An interrupted time-series analysis of preliminary data from 21 countries. *The Lancet Psychiatry 8*(7), 579–588.

Pleninger, R., S. Streicher, and J.-E. Sturm (2022). Do COVID-19 containment measures work? Evidence from Switzerland. *Swiss Journal of Economics and Statistics 158*(1), 1–24.

Praet, P. (2013). Forward guidance and the ECB. *VoxEU.org*.

Quast, J. and M. Wolters (2022). Reliable real-time output gap estimates based on a modified Hamilton filter. *Journal of Business and Economic Statistics 40*(1), 152–168.

Ravn, M. and H. Uhlig (2002). On adjusting the Hodrick-Prescott filter for the frequency of observations. *Review of Economics and Statistics 84*(2), 371–376.

Richter, D., S. Riedel-Heller, and S. Zuercher (2021). Mental health problems in the general population during and after the first lockdown phase due to the SARS-Cov-2 pandemic: Rapid review of multi-wave studies. *Epidemiology and Psychiatric Sciences*, 1–17.

Riecher-Rössler, A. (2017). Sex and gender differences in mental disorders. *The Lancet Psychiatry 4*(1), 8–9.

Ritchie, H., E. Mathieu, L. Rodés-Guirao, C. Appel, C. Giattino, E. Ortiz-Ospina, J. Hasell, B. Macdonald, D. Beltekian, and M. Roser (2020). Coronavirus pandemic (COVID-19). *Our World in Data*. https://ourworldindata.org/coronavirus.

Rith-Najarian, L., M. Boustani, and B. Chorpita (2019). A systematic review of prevention programs targeting depression, anxiety, and stress in university students. *Journal of Affective Disorders 257*, 568–584.

Rossi, R., V. Socci, D. Talevi, S. Mensi, C. Niolu, F. Pacitti, A. Di Marco, A. Rossi, A. Siracusano, and G. Di Lorenzo (2020). COVID-19 pandemic and lockdown measures impact on mental health among the general population in Italy. *Frontiers in Psychiatry*, 790.

Rubio-Ramirez, J., D. Waggoner, and T. Zha (2010). Structural vector autoregressions: Theory of identification and algorithms for inference. *The Review of Economic Studies 77*(2), 665–696.

Serrano-Alarcón, M., A. Kentikelenis, M. Mckee, and D. Stuckler (2022). Impact of COVID-19 lockdowns on mental health: Evidence from a quasi-natural experiment in England and Scotland. *Health Economics 31*(2), 284–296.

Silva, S. and S. Tenreyro (2006). The log of gravity. *The Review of Economics and Statistics 88*(4), 641–658.

Silverio-Murillo, A., L. Hoehn-Velasco, A. Tirado, and J. de la Miyar (2021). COVID-19 blues: Lockdowns and mental health-related Google searches in Latin America. *Social Science and Medicine 281*, 114040.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 1–48.

Slacalek, J. (2009). What drives personal consumption? The role of housing and financial wealth. *The BE Journal of Macroeconomics 9*(1).

Smets, F. (2002). Output gap uncertainty: Does it matter for the Taylor rule? *Empirical Economics 27*(1), 113–129.

Sorensen, R., R. Barber, D. Pigott, A. Carter, C. Spencer, S. Ostroff, R. Reiner Jr, C. Abbafati, C. Adolph, A. Allorant, et al. (2022). Variation in the COVID-19 infection-fatality ratio by age, time, and geography during the pre-vaccine era: A systematic analysis. *The Lancet 399*(10334), 1469–1488.

St-Amant, P. and S. van Norden (1997). Measurement of the output gap: A discussion of recent research at the Bank of Canada. *Technical Reports* (79).

Stock, J. and M. Watson (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting 23*(6), 405–430.

Stock, J. and M. Watson (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking 39*(1), 3–33.

Stock, J. and M. Watson (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In J. Taylor and H. Uhlig (Eds.), *Handbook of Macroeconomics*, Volume 2, Chapter 8, pp. 415–525. Elsevier.

Sturm, J.-E. and J. De Haan (2011). Does central bank communication really lead to better forecasts of policy decisions? New evidence based on a Taylor rule model for the ECB. *Review of World Economics 147*, 41–58.

Swanson, E. (2011). Let's twist again: A high-frequency event-study analysis of operation twist and its implications for QE2. *Brookings Papers on Economic Activity 2011*(1), 151–188.

Swanson, E. (2017). Measuring the effects of Federal Reserve forward guidance and asset purchases on financial markets. *NBER Working Paper* (23311).

Tanaka, T. and S. Okamoto (2021). Increase in suicide following an initial decline during the COVID-19 pandemic in Japan. *Nature Human Behaviour 5*(2), 229–238.

Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, pp. 135–196. Elsevier.

Turkington, R., M. Mulvenna, R. Bond, E. Ennis, C. Potts, C. Moore, L. Hamra, J. Morrissey, M. Isaksen, E. Scowcroft, et al. (2020). Behavior of callers to a crisis helpline before and during the COVID-19 pandemic: Quantitative data analysis. *JMIR Mental Health 7*(11), e22984.

Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics 52*(2), 381–419.

Ullrich, K. (2008). Inflation expectations of experts and ECB communication. *The North American Journal of Economics and Finance 19*(1), 93–108.

Van Der Cruijsen, C. and M. Demertzis (2007). The impact of central bank transparency on inflation expectations. *European Journal of Political Economy 23*(1), 51–66.

Vegh, C. and G. Vuletin (2015). How is tax policy conducted over the business cycle? *American Economic Journal: Economic Policy 7*(3), 327–370.

Vloo, A., R. Alessie, J. Mierau, M. Boezen, J. Mierau, L. Franke, J. Dekens, P. Deelen, P. Lanting, J. Vonk, et al. (2021). Gender differences in the mental health impact of the COVID-19 lockdown: Longitudinal evidence from the Netherlands. *SSM-Population Health 15*, 100878.

Wang, Y., Y. Di, J. Ye, and W. Wei (2021). Study on the public psychological states and its related factors during the outbreak of coronavirus disease 2019 (COVID-19) in some regions of China. *Psychology, Health and Medicine 26*(1), 13–22.

Watson, M. (1986). Univariate detrending methods with stochastic trends. *Journal of Monetary Economics 18*(1), 49–75.

Weathers, R. and M. Stegman (2012). The effect of expanding access to health insurance on the health and mortality of social security disability insurance beneficiaries. *Journal of Health Economics 31*(6), 863–875.

Wickramasuriya, S., G. Athanasopoulos, and R. Hyndman (2019). Forecasting hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association 114*(526), 804–819.

Wilson, J., J. Lee, and N. Shook (2021). COVID-19 worries and mental health: The moderating effect of age. *Aging and Mental Health 25*(7), 1289–1296.

Woodford, M. (2001). Monetary policy in the information economy. In *Economic Policy for the Information Economy*, pp. 297–370. Federal Reserve Bank of Kansas City.

Wu, J. C. and F. D. Xia (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit and Banking 48*(2-3), 253–291.

Zalsman, G., Y. Levy, E. Sommerfeld, A. Segal, D. Assa, L. Ben-Dayan, A. Valevski, and J. Mann (2021). Suicide-related calls to a national crisis chat hotline service during the COVID-19 pandemic and lockdown. *Journal of Psychiatric Research 139*, 193–196.

Zamarro, G. and M. J. Prados (2021). Gender differences in couples' division of childcare, work and mental health during COVID-19. *Review of Economics of the Household 19*(1), 11–40.

Curriculum Vitae

Marc Anderes

Address	Leonhardstrasse 21, 8092 Zurich, Switzerland
Email & Phone	anderes@kof.ethz.ch, +41 44 632 75 48
Birth Date & Place	January 20, 1990 in Aarau, Switzerland
Citizenship	Swiss

Research Interests

Economic Forecasting, Health Economics, Business Cycle Research, Bayesian Econometrics, Dynamic Linear Models

Current Position

Ph.D. Researcher in Economics, since September 2018
KOF Swiss Economic Institute, ETH Zurich, Switzerland
Supervisor: Jan-Egbert Sturm

Education

M.Sc. in International and Monetary Economics, University of Bern, July 2018
B.A. in Economics, University of St. Gallen, September 2015

B.A. in International Affairs, University of St. Gallen, September 2015

Refereed Journal Publications

Anderes, M., A. Rathke, S. Streicher, and J.E.-Sturm (2021). The role of ECB communication in guiding markets. *Public Choice 186*(3-4), 351-383.

Working Papers

Anderes, M. (2021). Housing Demand Shocks and Households' Balance Sheets. *KOF Working Papers*, vol. 492, Zurich: KOF Swiss Economic Institute, ETH Zurich.

Teaching

2021-2022: Teaching Assistant Time Series Econometrics and Macroeconomic Forecasting
Samad Sarferaz, ETH Zurich

Referee activities

European Journal of Political Economy.

Languages

German (native), English (fluent), French (intermediate)

Computational Skills

R, Matlab, Git, SQL, LaTex

KOF

ETH Zurich
KOF Swiss Economic Institute
LEE G 116
Leonhardstrasse 21
8092 Zurich, Switzerland

Phone +41 44 632 42 39
kof@kof.ethz.ch
www.kof.ch