DE GRUYTER Journal of Causal Inference 2021; 9: 109–146

Research Article

Philipp F. M. Baumann, Michael Schomaker*, and Enzo Rossi

# Estimating the effect of central bank independence on inflation using longitudinal targeted maximum likelihood estimation[1]

https://doi.org/10.1515/jci-2020-0016
received July 28, 2020; accepted May 15, 2021

**Abstract:** The notion that an independent central bank reduces a country's inflation is a controversial hypothesis. To date, it has not been possible to satisfactorily answer this question because the complex macroeconomic structure that gives rise to the data has not been adequately incorporated into statistical analyses. We develop a causal model that summarizes the economic process of inflation. Based on this causal model and recent data, we discuss and identify the assumptions under which the effect of central bank independence on inflation can be identified and estimated. Given these and alternative assumptions, we estimate this effect using modern doubly robust effect estimators, i.e., longitudinal targeted maximum likelihood estimators. The estimation procedure incorporates machine learning algorithms and is tailored to address the challenges associated with complex longitudinal macroeconomic data. We do not find strong support for the hypothesis that having an independent central bank for a long period of time necessarily lowers inflation. Simulation studies evaluate the sensitivity of the proposed methods in complex settings when certain assumptions are violated and highlight the importance of working with appropriate learning algorithms for estimation.

**Keywords:** causal inference, doubly robust, super learning, macroeconomics, monetary policy

**MSC 2020:** 62P20

## 1 Introduction

The impact of the institutional design of central banks on real economic outcomes has received considerable attention over the past three decades. Whether central bank independence (CBI) can lower inflation and provide inflation stability in a country is a particularly controversial issue. It has been claimed that more than 9,000 works have been devoted to the investigation of the role of CBI in influencing economic

1 The views, opinions, findings, and conclusions or recommendations expressed in this paper are strictly those of the authors. They do not necessarily reflect the views of the Swiss National Bank (SNB). The SNB takes no responsibility for any errors or omissions in, or for the correctness of, the information contained in this paper.

* **Corresponding author: Michael Schomaker,** Institute of Public Health, Medical Decision Making and Health Technology Assessment, Department of Public Health, Health Services Research and Health Technology Assessment, UMIT - University for Health Sciences, Medical Informatics and Technology, Hall in Tirol, Austria; Centre for Infectious Disease Epidemiology & Research, School of Public Health and Family Medicine, University of Cape Town, Cape Town, South Africa, e-mail: michael.schomaker@uct.ac.za
**Philipp F. M. Baumann:** KOF Swiss Economic Institute, Department of Management, Technology, and Economics (D-MTEC), ETH Zurich, 8092 Zurich, Zurich, Switzerland, e-mail: baumann@kof.ethz.ch
**Enzo Rossi:** Swiss National Bank, Department I, 8022 Zurich, Zurich, Switzerland; Department of Economics, University of Zurich, 8032 Zurich, Zurich, Switzerland, e-mail: enzo.rossi@snb.ch

outcomes [1]. After the 2008–2009 Global Financial Crisis, the debate on the optimal design of monetary policy authorities has become even more intense.

The statistical and economic literature is rich in studies that evaluate the relationship between CBI and inflation. A common approach is to treat countries as units in a linear regression model where inflation (the percentage change in the consumer price index [CPI]) is the outcome and a binary CBI index and several economic and political variables are covariates. While many studies have found that an independent central bank may lower inflation [2–7], other studies that have used a broader range of characteristics of a nation's economy have been unable to find such a relationship [8–10]. Moreover, there have been studies suggesting that the effect of CBI on inflation can only be seen during specific time periods [5] or only in developed countries [6,11,12].

Numerous articles have pointed out the weaknesses that come with simple cross-sectional regression approaches when evaluating the effect of CBI on inflation. First, the problem at hand is longitudinal in nature, and only an appropriate panel setup may be suitable to estimate the (long-term) effect of CBI on inflation. Second, the question of interest is essentially causal: i.e., what (average) inflation would we observe in 10 years' time, if – from now on – each country's monetary institution had an independent central bank compared to the situation in which the central bank was not independent? However, not all cross-sectional regression approaches embed their analyses in a holistic causal framework.

Some more recent work has attempted to overcome at least parts of these problems. For example, Crowe and Meade [13,14] use a panel data setup with two time intervals, and Klomp and De Haan [6] work with a random coefficient panel model to answer the question of interest in a longitudinal setup. Other authors, e.g., Walsh [15], acknowledge not only that current CBI may cause future inflation but also that current inflation is possibly related to future CBI status. Several authors have thus tried to use instrumental variable approaches to estimate the effect of CBI on inflation within a causal framework, but have been unable to find strong instruments [14,16].

It is clear that evaluating the effect of CBI on inflation requires a longitudinal causal estimation approach. However, it has been shown repeatedly that standard regression approaches are typically not suitable to answer causal questions, particularly when the setup is longitudinal and when the time-dependent confounders of the outcome–intervention relationship are affected by previous intervention decisions [17,18]. There are at least three methods to evaluate the effect of longitudinal (multiple time-point) interventions on an outcome in such complex situations: (1) inverse probability of treatment weighted (IPTW) approaches [19]; (2) standardization with respect to the time-dependent confounders (i.e., g-formula-type approaches [20,21]); and (3) doubly robust methods, such as targeted maximum-likelihood estimation (TMLE, see ref. [22]), which can be seen as a combination and generalization of the other two approaches.

Longitudinal targeted maximum likelihood estimation (LTMLE, see ref. [23]) is a doubly robust estimation technique that requires iteratively fitting models for the outcome and intervention mechanisms at each time point. With LTMLE, the causal quantity of interest (such as differences in counterfactual outcomes after intervening at multiple time points) is estimated consistently if either the iterated outcome regressions or the intervention mechanisms are estimated consistently. LTMLE, like other doubly robust methods, has an advantage over other approaches in that it can more readily incorporate machine learning methods while retaining valid statistical inference. Recent research has shown that this is important if correct model specification is difficult, such as when dealing with complex longitudinal data, potentially of small sample size, where relationships and interactions are most likely highly nonlinear and where the number of variables is large compared to the sample size [24,25].

Using causal inference in economics has a long history, starting with path analyses and potential outcome language [26,27] and continuing with regression discontinuity analyses [28], instrumental variable designs [29], and propensity score approaches in the context of the potential outcome framework [30], among many other methods. More recently, there have been works advocating the use of doubly robust techniques in econometrics [31]. From the perspective of statistical inference, this is a very promising suggestion because the integration of modern machine learning methods in causal effect estimation is almost inevitable in areas with a large number of covariates and complex data-generating processes [25].

However, the application of doubly robust effect estimation can be challenging for (macro-)economic data. First, the causal model that summarizes the knowledge about the data-generating process is often more complex for economic than for epidemiological questions, where most successful implementations have been demonstrated thus far [25,32–38]. The task of representing the causal model in a directed acyclic graph (DAG) becomes particularly challenging when considering how economic variables interact with each other over time. Thus, to build a DAG, a thorough review of a vast amount of literature is needed, and economic feedback loops need to be incorporated appropriately. Imbens [39], who discusses different schools of causal inference and their use in statistics and econometrics, as well as different estimation techniques, emphasizes this point:

> [⋯] a major challenge in causal inference is coming up with the causal model.

Second, even if a causal model has been developed, identification of an estimand has been established and data have been collected, statistical estimation may be nontrivial given the complexity of a particular data set [25]. If the sample size is small, potentially smaller than the number of (time-varying) covariates, recommended estimation techniques can fail, and the development of an appropriate set of learning and screening algorithms is important. The benefits of LTMLE, which is doubly robust effect estimation in conjunction with machine learning to reduce the chance of model misspecification, can be best utilized under a good and broad selection of learners that are tailored to the problem of interest.

Estimating the effect of CBI on inflation is a typical example of a causal inference question that faces all of the challenges described above. Our article makes five novel contributions to the literature. (i) We discuss identification and estimation for our question of interest and estimate the effect of CBI on inflation; (ii) we develop a causal model that can be applied to other macroeconomic questions; (iii) we demonstrate that it is possible to develop a DAG for economic questions, which is important, as it has been argued that "*the lack of adoption in economics is that the DAG literature has not shown much evidence of the benefits for empirical practice in settings that are important in economics*." [39]; (iv) we demonstrate how to integrate machine learning into complex causal effect estimation, including how to define a successful learner set when the number of covariates is larger than the sample size and when there is time-dependent confounding with treatment-confounder feedback [40]; and (v) we use simulation studies to study the performance of doubly robust estimation techniques under the challenges described above.

This article is structured as follows. In Section 2, we motivate our question of interest, and the general description of our framework is given in Section 3. Section 4 contains the data analysis and describes the doubly robust estimation strategy to estimate the effect of CBI on inflation. In Section 5, we conduct simulation studies motivated by our data analysis. Section 6 concludes the article.

## 2 Motivating question: CBI and inflation

When governments have discretionary control over monetary instruments, typically a short-term interest rate, they can prioritize other policy goals over price stability. For instance, after nominal wages have been negotiated (or nominal bonds purchased), politicians may be tempted to create inflation to boost employment and output (gross domestic product [GDP]) or to devalue government debt. This is referred to as the time-inconsistency problem of commitments to price stability. It results in an inflation rate higher than what is socially desirable. To overcome this outcome, the literature discusses a variety of commitment mechanisms (also called "commitment technologies"), ranging from simple rules (such as the imposition of strict rules on the rate of monetary expansion, inflation targeting and nominal exchange rate targeting), contracts between the government and the central bank, reputational forces and, from a practical perspective the best known and implemented mechanism, the delegation of monetary policy-making to an independent central bank. In particular, Rogoff [41] has proposed delegating monetary policy to an independent and "conservative" central banker to reduce the tendency to produce high inflation. Here, conservative means that the central banker dislikes inflation more than the government, in the sense that they places a greater weight on price stability than the government does. Once central bankers are insulated from political pressures, commitments to price stability can be credible, which helps to maintain low inflation. Rogoff's seminal paper had a twofold effect: stimulating the implementation of central bank reforms on the policy side and creating avenues for the design of indices that are suitable to capture the degree of independence of these institutions on the research side.

Following these ideas, a considerable policy consensus grew around the potential of having independent central banks to promote inflation stability [42,43]. Numerous countries followed this policy advice. Between 1985 and 2012, and excluding the creation of regional central banks, there were 266 reforms to the statutory independence of central banks, 236 of which were being implemented in developing countries. Most of these reforms (77%) strengthened CBI [44], though some also weakened it. For instance, the law governing the Reserve Bank of Australia was changed in 2002. While previously the governor and board members were appointed by the governor general; in 2002 appointment power was given to the treasurer, which produced a lower independence score. Moreover, whereas board members had been appointed for exactly 5 years before this amendment, after the amendment the term was specified as not exceeding 5 years at the discretion of the appointing person [45].

Despite the broad impact of the policy advice to make central banks more independent, the empirical evidence in support of it remains controversial. We investigate the effect of CBI on inflation with a causal framework that treats countries as units in a longitudinal (panel) setup. The data set we use in our analysis was created specifically for this purpose and extends the data set from Baumann et al. [46]. To describe and address relevant confounding structures, the crucial question is: What are possible reasons that motivate the decision of a country to adopt a certain degree of CBI? What macroeconomic factors drive CBI [47–54]? Four arguments stand out:

(i) *Political institutions*: Federally organized countries with good checks and balances grant their monetary institutions greater autonomy and thus a greater level of CBI [47,55,56].

(ii) *Political instability*: Central bank reforms are more likely to follow elections, which lead to political consolidation or to changes in the political orientation of the government [53]. Cukierman and Webb [57] found that *de facto* CBI, as measured by the turnover rate of the central bank governor, is lower in less stable political systems.

(iii) *Past inflation*: Crowe and Meade [14] showed that over the period 1990–2003, greater changes in CBI have occurred in countries originally characterized by lower levels of independence and higher inflation. This finding is strengthened by the research of Masciandaro and Romelli [54], where it is shown that countries which experienced long periods of inflation are characterized by a higher inflation aversion, which may cause the government to grant a higher level of CBI. According to Wachtel and Blejer [58], the arguments in favor of an independent central bank began to crystallize in the 1980s after a decade or more of traumatic inflationary experience that put a spotlight on central bank policymaking and its failures.

(iv) *International pressure*: Binding agreements with international money lenders like the International Monetary Fund or the World Bank often require countries to commit to a particular set of policies [43,49,53,59–62]. According to Dincer and Eichengreen [45], countries with less developed financial markets, more open economies, and countries that have participated in IMF programs have more independent central banks. Similarly, Romelli [53] found that countries receiving an IMF loan or becoming a member of a currency union adopt reforms that increase CBI. Another type of external pressure can come from regional clustering, which is often found to be cohesive of certain types of reform processes such as democratizations and economic liberalizations [63–66].

Those arguments inform our causal model and estimation strategies in Section 4.

# 3 Methodological framework

## 3.1 Notation

We consider panel data with $n$ units (i.e., countries in our case) studied over time ($t = 0, 1, \ldots, T$). At each time point $t$, we observe an outcome $Y_t$, an intervention of interest $A_t$, and several time-dependent covariates $L_t^j$, $j = 1, \ldots, q$, collected in a set $\mathbf{L}_t = \{L_t^1, \ldots, L_t^q\}$. Variables measured at the first time point ($t = 0$) are denoted as $\mathbf{L}_0 = \{L_0^1, \ldots, L_0^{q_0}\}$ and are called "baseline variables." The intervention and covariate histories of a unit $i$ (up to and including time $t$) are $\bar{A}_{t,i} = (A_{0,i}, \ldots, A_{t,i})$ and $\bar{L}_{t,i}^s = (L_{0,i}^s, \ldots, L_{t,i}^s)$, $s = 1, \ldots, q$, respectively, with $q, q_0 \in \mathbb{N}$.

We are interested in the counterfactual outcome $Y_{t,i}^{\bar{a}_t}$ that would have been observed at time $t$ if unit $i \in \{1, \ldots, n\}$ had received, possibly contrary to the fact, the intervention history $\bar{A}_{t,i} = \bar{a}_t$. For a given intervention $\bar{A}_{t,i} = \bar{a}_t$, the counterfactual covariates are denoted as $\bar{\mathbf{L}}_{t,i}^{\bar{a}_t}$. If an intervention depends on covariates, it is dynamic. A dynamic intervention $d_t(\bar{\mathbf{L}}_t) = \bar{d}_t$ assigns treatment $A_{t,i} \in \{0, 1\}$ as a function $\bar{\mathbf{L}}_{t,i}$. If $\bar{\mathbf{L}}_{t,i}$ is the empty set, the treatment $\bar{d}_t$ is static. We use the notation $\bar{A}_t = \bar{d}_t$ to refer to the intervention history up to and including time $t$ for a given rule $\bar{d}_t$. The counterfactual outcome at time $t$ related to a dynamic rule $\bar{d}_t$ is $Y_{t,i}^{\bar{d}_t}$, and the counterfactual covariates at the respective time point are $\bar{\mathbf{L}}_{t,i}^{\bar{d}_t}$. More specific notation concerning the data analysis is given in Section 4.

## 3.2 Likelihood

If we assume a time ordering of $\mathbf{L}_t \rightarrow A_t$ at each time point, use $Y_T$ as the outcome, and define $Y_t$, $t < T$, to be contained in $\mathbf{L}_t$, the data can be represented as $n$ iid copies of the following longitudinal data structure:

$$O = (\mathbf{L}_0, A_0, \mathbf{L}_1, A_1, \ldots, \mathbf{L}_{T-1}, A_{T-1}, Y_T) \overset{iid}{\sim} P_0.$$

Note that in Section 4.3, in the data analysis, the ordering of variables is different. However, for the given ordering, we can write the respective likelihood $\mathcal{L}(O)$ as

$$\begin{aligned} p_0(O_i) &= p_0(\mathbf{L}_{0,i}, A_{0,i}, \mathbf{L}_{1,i}, A_{1,i}, \ldots, \mathbf{L}_{T-1,i}, A_{T-1,i}, Y_{T,i}) \\ &= p_0(Y_{T,i} | \bar{A}_{T-1,i}, \bar{\mathbf{L}}_{T-1,i}) \times p_0(A_{T-1} | \bar{\mathbf{L}}_{T-1,i}, \bar{A}_{T-2,i}) \times p_0(\mathbf{L}_{T-1} | \bar{A}_{T-2,i}, \bar{\mathbf{L}}_{T-2,i}) \times \ldots \times p_0(\mathbf{L}_{0,i}) \\ &= p_0(Y_{T,i} | \bar{A}_{T-1,i}, \bar{\mathbf{L}}_{T-1,i}) \left[ \prod_{t=0}^{T-1} \underbrace{p_0(A_{t,i} | \bar{\mathbf{L}}_{t,i}, \bar{A}_{t-1,i})}_{g_{0,A_t}} \right] \times \left[ \prod_{t=0}^{T-1} \underbrace{p_0(\mathbf{L}_{t,i} | \bar{A}_{t-1,i}, \bar{\mathbf{L}}_{t-1,i})}_{\tilde{q}_{0,\mathbf{L}_t}} \right]. \end{aligned}$$

In the above factorization, $p_0(\cdot)$ refers to the density of $P_0$ (with respect to some dominating measures) and $A_{-1} := \mathbf{L}_{-1} := \emptyset$. If an order for $\mathbf{L}_t$ is given, e.g., $L_t^1 \rightarrow \cdots \rightarrow L_t^q$, a more refined factorization is possible. In line with the notation of other papers (e.g., ref. [24]), we define the *q-portion* of the likelihood to also contain the outcome: $q_{0,\mathbf{L}_t} := \tilde{q}_{0,\mathbf{L}_t} \times p_0(Y_{T,i} | \bar{A}_{T-1,i}, \bar{\mathbf{L}}_{T-1,i})$. Similarly, we define $g_0 := \prod_{t=0}^{T} g_{0,A_t}$ and $q_0 := \prod_{t=0}^{T} q_{0,\mathbf{L}_t}$.

## 3.3 On the distinction between the causal and statistical model

Estimating causal effects cannot be established from data alone but requires additional structural (i.e., causal) assumptions about the data-generating process. Therefore, any causal analysis comes with both a structural (i.e., causal) and a statistical model. The former can be represented by a DAG, which encodes conditional independence assumptions and is logically equivalent to a (non-parametric) structural equation framework. Ideally, the structural model is supported by knowledge from the literature. The statistical model encodes assumptions about the family of possible observed data distributions associated with the DAG, with the ultimate aim to estimate post-intervention distributions and quantities. With doubly robust effect estimation, any parametric assumptions are typically eschewed to avoid model mis-specification and to incorporate machine learning while retaining valid inference. In our framework and analyses below, we proceed as follows: for the causal model, we begin with the basic assumption that variables can be affected by the past, but not the future (Section 3.5). In our analysis in Section 4.1, we then make more detailed assumptions with respect to the causal model: we encode our structural assumptions in a DAG (Figure 1) and support this model with references from the economic literature (Appendix). For the statistical model, we first do not impose any parametric restrictions on the statistical model (Section 3.4). In the analysis (Section 4.1), we then use the above likelihood factorization and TMLE with super learning, to avoid any overly restrictive parametric assumptions.

## 3.4 Statistical model

In line with the notation of Section 3.2, we consider a statistical model $\mathcal{M} = \{P = q \times g : q \in \mathcal{Q}, g \in \mathcal{G}\}$ for the true distribution $P_0$ that requires minimal (parametric) assumptions. In contrast to many medical applications, we do not impose restrictions on this model; that is, $A_t$ and $Y_t$ are not deterministically determined for any given data history. Once an intervention is implemented, it can be stopped at any time point and potentially started again. Similarly, the outcome can be observed at any time point, and we do not assume that censoring is possible.

## 3.5 Causal model

Causal assumptions about the data-generating process are encoded in the model $\mathcal{M}^{\mathcal{F}}$. This nonparametric (structural equation) model states our assumptions about the time ordering of the data and the causal mechanism that gave rise to the data. Thus far, it relates to

$$\begin{aligned} Y_T &= f_{Y_T}\left(\bar{A}_{T-1}, \bar{\mathbf{L}}_{T-1}, U_{Y_T}\right), \\ \boldsymbol{L_t} &= f_{\boldsymbol{L_t}}\left(\bar{A}_{t-1}, \bar{\boldsymbol{L}}_{t-1}, \boldsymbol{U_{L_t}}\right) : t = 0, 1, \ldots, T-1, \\ A_t &= f_{A_t}\left(\bar{\boldsymbol{L}}_t, \bar{A}_{t-1}, U_{A_t}\right) : t = 0, 1, \ldots, T-1, \end{aligned}$$

where $\mathbf{U} := (U_{Y_T}, \boldsymbol{U_{L_t}}, U_{A_t})$ are unmeasured variables from some underlying distribution $P_{\boldsymbol{U}}$. For now, we do not make any assumptions regarding $P_{\boldsymbol{U}}$. However, in the data example further below, we need to enforce some restrictions on this distribution. The functions $f_O(\cdot)$ are (deterministic) nonparametric structural equations that assume that each variable may be affected only by variables measured in the past and not those that are measured in the future. Section 4.3 refines the causal model for the data-generating process of the motivating question and represents any additional assumptions made in a DAG.

## 3.6 Causal target parameter and identifiability

In this article, we focus on the differences in intervention-specific means, i.e., in target parameters such as

$$\psi_{j,k} = \mathbb{E}\left(Y_T^{\bar{d}_t^j}\right) - \mathbb{E}\left(Y_T^{\bar{d}_t^k}\right), \quad j \neq k. \tag{1}$$

If we set the intervention according to a static or dynamic rule ($\bar{A}_t = \bar{d}_t^l\ \forall t$) with $l \in \{j, k\}$ in the causal model $\mathcal{M}^{\mathcal{F}}$, we obtain the post-intervention distribution $P_0^{\bar{d}_t^l}$. The counterfactual outcome $Y_T^{\bar{d}_t^l}$ is defined as the outcome that would have been observed had $A_t$ been set deterministically to 0 or 1 according to rule $\bar{d}_t^l$. We thus restrict the set of possible interventions to those where the intervention is binary $A_{t,i} \in \{0, 1\}$.

It has been shown that target parameters of the form 1 can be identified under the (partly untestable) assumptions of consistency, conditional exchangeability, and positivity, which are defined below. Specifically, it follows from the work of Bang and Robins [21] that given these three assumptions, using the iterative conditional expectation rule, and for the particular time-ordering as defined in Section 3.2, we can write the target parameter as

$$\begin{aligned}\psi_{j,k} &= \mathbb{E}\left(Y_T^{\bar{d}_t^j}\right) - \mathbb{E}\left(Y_T^{\bar{d}_t^k}\right)\\ &= \mathbb{E}(\mathbb{E}(\mathbb{E}(\ldots\mathbb{E}(\mathbb{E}(Y_T|\bar{\mathbf{A}}_{T-1} = \bar{d}_{T-1}^j, \bar{\mathbf{L}}_{T-1})|\bar{\mathbf{A}}_{T-2} = \bar{d}_{T-2}^j, \bar{\mathbf{L}}_{T-2})\ldots|\bar{A}_0 = \bar{d}_0^j, \mathbf{L}_0)|\mathbf{L}_0)\\ &\quad - \mathbb{E}(\mathbb{E}(\mathbb{E}(\ldots\mathbb{E}(\mathbb{E}(Y_T|\bar{\mathbf{A}}_{T-1} = \bar{d}_{T-1}^k, \bar{\mathbf{L}}_{T-1})|\bar{\mathbf{A}}_{T-2} = \bar{d}_{T-2}^k, \bar{\mathbf{L}}_{T-2})\ldots|\bar{A}_0 = \bar{d}_0^k, \mathbf{L}_0)|\mathbf{L}_0).\end{aligned} \tag{2}$$

The assumptions of consistency, conditional exchangeability, and positivity have been discussed in the literature in detail [18,24,67–69]. Briefly, consistency is the requirement that $Y_T^{\bar{d}_t} = Y_T$ if $\bar{\mathbf{A}}_{t-1} = \bar{d}_{t-1}$ and $\bar{\mathbf{L}}_t^{\bar{d}_t} = \bar{\mathbf{L}}_t$ if $\bar{\mathbf{A}}_{t-1} = \bar{d}_{t-1}$. Conditional exchangeability requires the counterfactual outcome under the assigned treatment rule to be independent of the observed treatment assignment, given the observed past: $Y_T^{\bar{d}_t} \coprod \mathbf{A}_{t-1}|\bar{\mathbf{L}}_{t-1}, \bar{\mathbf{A}}_{t-2}\ \forall \bar{\mathbf{A}}_t = \bar{d}_t,\ \bar{\mathbf{L}}_t = \bar{\mathbf{l}}_t,\ \forall t$, and positivity says that each unit should have a positive probability of continuing to receive the intervention according to the assigned treatment rule, given that this has been done so far, and irrespective of the covariate history: $P(\mathbf{A}_t = \bar{d}_t|\bar{\mathbf{L}}_t = \bar{\mathbf{l}}_t, \bar{\mathbf{A}}_{t-1} = \bar{d}_{t-1}) > 0\ \forall t, \bar{d}_t, \bar{\mathbf{l}}_t$ with $P(\bar{\mathbf{L}}_t = \bar{l}_t, \bar{\mathbf{A}}_{t-1} = \bar{d}_{t-1}) \neq 0$.

In principle, (conditional) exchangeability can be evaluated graphically for an assumed structural model represented in a DAG using the back-door criterion [70,71]; i.e., by closing all back-door paths and by nonconditioning on descendants of the intervention. For multiple time-point interventions, a generalized version of this criterion can be used to verify conditional exchangeability. This requires blocking all back-door paths from $A_t$ to $Y_T$ that do not go through any future treatment node $A_{t+1}$ [40]. More generally, it has been suggested to use single-world intervention graphs to verify exchangeability, particularly to evaluate identification for complex dynamic interventions. See the study of Richardson and Robins for details [72].

## 3.7 Effect estimation with longitudinal TMLE

The longitudinal TMLE estimator [23] relies on equation (2). To estimate $\psi_{j,k}$, one can separately evaluate each of the two nested expectation terms and integrate out $\bar{\mathbf{L}}_{T-1}$ with respect to the post-intervention distribution $P_0^{\bar{d}_t^l}$. To improve inference with respect to $\psi_{j,k}$, a targeted estimation step at each time point yields a doubly robust estimator of the desired target quantity (see the study of Van der Laan and Rose [22] or Schnitzer and Cefalu [73] for details). Specifically, we recur to the following algorithm for $t = T, \ldots, 1$:

1. Estimate $\bar{Q}_T = \mathbb{E}(Y_T|\bar{\mathbf{A}}_{T-1}, \bar{\mathbf{L}}_{T-1})$ with an appropriate model (for $t = T$). If $t < T$, use the prediction from step 3d (of iteration $t-1$) as the outcome and fit the respective model. The estimated model is denoted as $\hat{Q}_{0,t}$.
2. Now, plug in $\bar{\mathbf{A}}_{t-1} = \bar{d}_{t-1}^l$ based on rule $\bar{d}_t^l$ and use the fitted model from step 1 to predict the outcome at time $t$ (which we denote as $\hat{Q}_{0,t}^{\bar{d}_t^l}$).
3. To improve estimation with respect to the target parameter, update the initial estimate of step 2 by means of the following regression:
   (a) The outcome refers again to the measured outcome for $t = T$ and to the prediction from item 3d (of iteration $t-1$) if $t < T$.
   (b) The offset is the original predicted outcome $\hat{Q}_{0,t}^{\bar{d}_t^l}$ from step 2 (iteration $t$).
   (c) The "clever covariate" is defined as:

$$H_{t-1} = \prod_{s=0}^{t-1} \frac{I(\bar{A}_s = \bar{d}_s)}{g_{0,A_t=\bar{d}_s^l}} \tag{3}$$

   with $g_{0,A_t=\bar{d}_s^l} = P(A_s = \bar{d}_s^l|\bar{\mathbf{L}} = \bar{\mathbf{l}}_s, \bar{A}_{s-1} = \bar{d}_{s-1}^l)$. The estimate of $g_{0,A_t=\bar{d}_s^l}$ is denoted as $\hat{g}_{A_t=\bar{d}_s^l}$.
   (d) Predict the updated (nested) outcome, $\hat{Q}_{1,t}^{\bar{d}_t^l}$, based on the model defined through 3a, 3b, and 3c.

This model contains no intercept. Alternatively, the same model can be fitted with $H_{t-1}$ as a weight rather than a covariate [24,32]. In this case, an intercept is required. We follow the latter approach in our implementations.

4. The estimate for $\mathbb{E}(Y_T^{\bar{d}_t^l})$ is obtained by calculating the mean of the predicted outcome from step 3d (where $t = 1$).
5. Confidence intervals can, for example, be obtained using the vector of the estimated influence curve; see the study of Tran et al. [74] for a review of adequate choices.
6. Repeat 1–5 to estimate $\mathbb{E}(Y_T^{\bar{d}_t^j})$ and $\mathbb{E}(Y_T^{\bar{d}_t^k})$. Now, $\hat{\psi}_{j,k}$ and its corresponding confidence intervals can be calculated.

### 3.7.1 Inference and properties of LTMLE

For an arbitrary distribution $P \in \mathcal{M}$ and a specific intervention rule $g = g(P)$, we consider the statistical model $M(g) = \{P^* \in \mathcal{M} : g(P^*) = g\}$ for the respective treatment rules $g$. With such a model we could estimate $\psi^*$ with the algorithm described in 3.7. For $\psi^*$ it can be shown (e.g., ref. [75]) that $\hat{\psi}^*$ is an asymptotically efficient estimator of $\psi^*$ where

$$\sqrt{n}(\hat{\psi}^* - \psi^*) \xrightarrow{d} N(0, \sigma^{2,*}). \tag{4}$$

The variance can be estimated with the sample variance of the estimated influence curve. This is essentially because the construction of the covariate in step 3c, guarantees that the estimating equation corresponding to the (efficient) influence curve is solved, which in turn yields desirable (asymptotic) inferential properties. The influence curve emerges from the linear span of the scores (i.e., first derivative) of the logistic loss for the density of the outcome variable (evaluated at zero) for a given value of the clever covariate [35]. Thus, in the longitudinal case, for interventions rules $\bar{g}_t$, these score components can be summed across the points in time which yields the efficient influence curve

$$\hat{IC}^* = \left\{ \sum_{t=1}^{T} \hat{H}_{t-1}^* \left[ \hat{Y}_t^{*\bar{d}_t = \bar{g}_t} - \hat{Y}_{t-1}^{*\bar{d}_{t-1} = \bar{g}_{t-1}} \right] \right\} + \hat{Y}_0^{*\bar{d}_t = \bar{g}_t} - \hat{\psi}^*. \tag{5}$$

### 3.7.2 Data-adaptive estimation for complex (macroeconomic) data

The above estimation procedure is doubly robust, which means that the estimator is consistent as long as either the Q- or the g-models (steps 1 and 3c in the algorithm described above) are estimated consistently [21]. If both are estimated consistently (at reasonable rates), the estimator is asymptotically efficient because the construction of the covariate in step 3c guarantees that the estimating equation corresponding to the efficient influence curve is solved, which in turn yields desirable (asymptotic) inferential properties [22,73].

To estimate the conditional expectations in the algorithm, one could use (parametric) regression models. Under the assumption that they are correctly specified, this approach would be valid. However, in the context of complex macroeconomic data, as in our motivating example below, it is challenging to estimate appropriate parametric models because of small sample sizes, a large number of relevant variables and complex nonlinear relationships. Longitudinal TMLE can (in contrast to many competing estimation techniques) incorporate machine learning algorithms while still retaining valid inference to reduce the possibility of model misspecification. However, in the settings presented below, machine learning approaches need to be tailored to the specific problem and address the following challenges:

(i) **Complexity**: Macroeconomic relationships are often highly nonlinear and have various interactions of higher order, which need to be modeled in a sophisticated manner while taking into account the time ordering of the data.

(ii) **Dispensable variables**: The inclusion of covariates in the estimation procedure that are not required for identification, i.e., do not block any back-door paths, can potentially be harmful even if they are not colliders or mediators [76]; that is, the inclusion of such variables can increase the finite-sample variance and lead to small estimated probabilities of following a particular treatment rule given the past, which may be both incorrectly interpreted as positivity violations and make the updating step in the TMLE algorithm unstable. They may also amplify bias [77].

(iii) $\mathbf{p} > \mathbf{n}$: For longitudinal macroeconomic data, the number of parameters is often larger than the sample size. This is because for long follow-up, the whole covariate history needs to be considered, interactions may be nonlinear, and different variables may have different scales and features that need to be modeled adequately. Consequently, one needs to either reduce the number of parameters with an appropriate estimation procedure or eliminate variables beforehand using variable screening. It has been argued that screening of variables is inevitable to facilitate estimation with LTMLE in many settings [76].

Section 4.5 recommends possible approaches to tackle these challenges in common macroeconomic settings.

# 4 Data analysis: estimating the effect of CBI on inflation

## 4.1 Data

We accessed databases of the World Bank and the International Monetary Fund to collect annual data for economic, political, and institutional variables. Our outcome of interest is inflation in 2010 ($Y_{2010}$). All covariates are measured annually at equidistant points in time for $t^* = 1998, \ldots, 2010$. The intervention variable is CBI at time $t^*$ (CBI, $A_{t^*}$), which we define as suggested by Dincer and Eichengreen [45]: their CBI index measures several dimensions of independence and runs from 0, the lowest level of independence, to 1, the highest level of independence. It contains considerations such as the independence of the chief executive officer (CEO) and limits on his/her reappointment, the bank's independence in terms of policy formulation, its objective or mandate, the stringency of limits on lending money to the public sector, measures of provisions affecting (re)appointment of board members other than the CEO, restrictions on government representation on the board, and intervention of the government in exchange rate policy formulation. This definition implies that our CBI index is an intervention that can in principle be modified through legislative amendments, although it is the very nature of an index to represent multiple facets of a phenomenon that cannot be easily dealt with in an actual experiment. We binarized the index of Dincer and Eichengreen [45] at a value of 0.45 by setting countries with a value greater than 0.45 to 1 (independent) for each time point and 0 (dependent) otherwise. We then used the binarized index for estimation. The trajectories of their original indices and our binarized version can be seen in Figure 6. Our outcome variable is defined as the year-on-year changes (expressed as annual percentages) of average consumer prices measured by a CPI. A CPI measures changes in the prices of goods and services that households consume. To calculate CPIs, government agencies conduct household surveys to identify a basket of commonly purchased items and then track the cost of purchasing this basket over time. The cost of this basket at a given time, expressed relative to a base year, is the CPI, and the percentage change in the CPI over a certain period is referred to as consumer price inflation, the most widely used measure of inflation. Our measured covariates are $\mathbf{L}_{t^*} = \{L_{t^*}^1, \ldots, L_{t^*}^{18}\}$ and include a variety of macroeconomic variables such as money supply, energy prices, economic openness, institutional variables such as central bank transparency and monetary policy strategies, and political variables (see Figure 1, Table 2 and Baumann et al. [46] for details). In line with the notation of Section 3, we consider $Y_{t^*}$, $t^* < T = 2010$, to be part of $\mathbf{L}_{t^*}$, i.e., we define $L_{t^*}^8 := Y_{t^*}$.

Our aim was to include as many countries as possible in our analysis. This entailed a tradeoff between the number of countries and the completeness of the data set. We were able to collect annual data from 1998 to 2010 for 124 countries for 14 explanatory variables and for the dependent variable $Y_{t^*}$. We further derived growth rates and other indicators from those measured variables to capture data for all 18 covariates ($\mathbf{L}_{t^*}$).

Some of the data were missing, however. To decide whether the missing data were likely missing not at random (MNAR) and therefore possibly not useful without making additional assumptions, we examined countries' characteristics. We decided that observations for certain variables, countries or groups of countries had to be excluded because they were not available; for instance, sometimes wars, insufficiently developed institutions, social unrest, or other reasons made the collection of data impossible. We split the data set according to our assessment of whether the observation was MNAR. Data that we regarded as missing at random (MAR) (2.7% of the data set) were multiply imputed using `Amelia II`[78], taking the time-series cross-sectional structure of the data into account. We did not impute data that were likely MNAR. However, some variables that were categorized as MNAR were used in the analysis (e.g., CBI). As a result, we obtained observations for 60 countries and 13 points in time (i.e., calendar years 1998–2010) for 19 measured variables ($L^1_{t^*}, \dots, L^7_{t^*}, L^9_{t^*}, \dots, L^{18}_{t^*}, Y_{t^*} \equiv L^8_{t^*}, A_{t^*}$). In this final data set, 0.1% of observations were missing and thus imputed.

According to the World Bank's income classification, approximately 20% of the remaining 60 countries are low-income countries, 36% belong to the lower-middle-income category, 27% to the upper-middle-income category, and 17% belong to the high-income category. While our sample reflects considerable heterogeneity with respect to countries' development level, it is possible that the included countries are not representative of all countries in the world: as many excluded countries faced periods of violent conflicts or had no well-developed governmental institutions, our sample likely reflects economies of (reasonably) stable countries.

## 4.2 Target parameters and interventions

Our target parameters are average treatment effects (ATEs) as defined in 1. To be more specific, consider the following three interventions, of which two are static and one dynamic, each of them applied to $\forall t^* \in \{1998, \dots, 2008\}$:

$$\bar{d}^1_{t^*} = \{a_{t^*} = 1$$

$$\bar{d}^2_{t^*,i}\left(\bar{L}^8_{t^*-1}\right) = \begin{cases} a_{t^*,i} = 1 & \text{if } \widehat{\text{median}}\left(L^8_{t^*-1,i}, \dots, L^8_{t^*-7,i}\right) \le 0 \text{ or } \widehat{\text{median}}\left(L^8_{t^*-1,i}, \dots, L^8_{t^*-7,i}\right) \ge 5 \\ a_{t^*,i} = 0 & \text{otherwise} \end{cases}$$

$$\bar{d}^3_{t^*} = \{a_{t^*} = 0.$$

A country's central bank is set to be either independent (i.e., $\bar{d}^1_{t^*}$) or dependent (i.e., $\bar{d}^3_{t^*}$) during the whole time period under the first and third intervention above. This means that we intervene on the first 11 (i.e., from 1998 to 2008) out of 13 points (i.e., from 1998 to 2010) in time. This is because we assume a 2-year lag between the CBI intervention and its effect on inflation. The transmission mechanism of monetary policy is said to exhibit "long and variable" lags [79–81]. In line with this view, inflation-targeting central banks have adopted a value between 12 and 24 months as transmission lag (the horizon at which the response of prices becomes the strongest). Theoretical models usually imply transmission lags of similar length [82]. According to a meta analysis of 67 published studies for 30 different countries [83] the average transmission lag is 29 months. However, transmission lags are longer in developed economies (25–50 months) than in post-transition economies (10–20 months). Overall, after filtering out effects of misspecifications, the results suggest that prices bottom out approximately two-and-a-half years after a monetary contraction. Given the heterogeneity of our countries, we chose a somewhat shorter lag to take into account countries' differences in their stage of development. The second (dynamic) intervention sets a country's central bank to be independent if its median inflation rate in the past 7 years was below 0% or greater than 5%. The rationale for this relates to the fact that excessive inflation and deflation over several years are considered to produce harmful effects on a country's economy (see, e.g., Tobin [84], Fisher [85]). To guarantee price stability, which excludes inflation beyond a certain level and deflation, an independent central bank is required. Over the last 20 years, the optimal level of inflation has been associated with approximately 2% [86]. If a country's inflation is constantly well above this level, in our case 5%, it will change the status of its central bank towards, independence. The same holds for an inflation rate systematically falling below a value of zero. Note that for the dynamic intervention $\bar{d}^2_{t^*,i}$, data prior to 1998 had to be collected and utilized.

We define the following two target parameters:

$$\psi_{1,3} = \mathbb{E}\left(Y_{2010}^{\bar{d}^1_{t^*}}\right) - \mathbb{E}\left(Y_{2010}^{\bar{d}^3_{t^*}}\right), \tag{6}$$

$$\psi_{2,3} = \mathbb{E}\left(Y_{2010}^{\bar{d}^2_{t^*}}\right) - \mathbb{E}\left(Y_{2010}^{\bar{d}^3_{t^*}}\right). \tag{7}$$

The first, $\psi_{1,3}$, quantifies the expected difference in inflation 2 years after the last intervention (i.e., in 2010) if every country had an independent central bank for 11 years in a row compared to a dependent central bank for 11 consecutive years. The second, $\psi_{2,3}$, quantifies the effect that would have been observed if every country's central bank had become independent for time points when the country's median inflation in the preceding 7 years had been outside the range from 0 to 5, compared to a strictly dependent central bank for 11 consecutive years (i.e., for the period 1998–2008).

## 4.3 Statistical and causal model (DAG)

We separate the measured variables into blocks. The first block comprises $\mathbf{L}^A_{t^*} := \{L^1_{t^*}, \ldots, L^7_{t^*}, L^9_{t^*}, \ldots, L^{15}_{t^*}\}$, and the second comprises $\mathbf{L}^B_{t^*} := \{L^{16}_{t^*}, \ldots, L^{18}_{t^*}\}$. In line with Sections 3.2 and 3.4, we do not make any overly restrictive assumptions with respect to our statistical model. First, we assume that our data come from a general true distribution $P_0$ and are ordered such that

$$O = (Y_{1998}, \mathbf{L}^A_{1998}, A_{1998}, \mathbf{L}^B_{1998}, Y_{1999}, \mathbf{L}^A_{1999}, A_{1999}, \mathbf{L}^B_{1999} \ldots, Y_{2009}, \mathbf{L}^A_{2009}, A_{2009}, \mathbf{L}^B_{2009}, Y_{2010}) \overset{iid}{\sim} P_0.$$

In the context of our application, we do not need to make any deterministic assumptions regarding our intervention assignment: a central bank can, in principle, be independent or dependent at any point in time, irrespective of the country's history – and thus be intervened upon.

As discussed in Section 3.5, we assume that each variable may be affected only by variables measured in the past and not those that are measured in the future. In addition, we make several assumptions regarding the data-generating process, which are summarized in the DAG in Figure 1. Not all variables listed in $O$ are needed during estimation; see Section 4.5.

The DAG contains both measured variables (in grey color) and unmeasured variables (in white color). The outcome variable is colored in green, and the intervention in red.

The DAG summarizes our knowledge of the transmission channels of monetary policy. An arrow $A \to B$ reflects our belief, corroborated by economic theory, that $A$ may cause $B$, whereas an absence of such an arrow states that we assume no causal relationship between the respective two variables. Figure 1 has been developed based on economic theory. For example, arrow number 6 describes the causal effect from real GDP (Output) on one component of companies' price setting (Price Markup), which is motivated by the fact that changes in demand (c.p.) in the goods market enable companies to set higher prices in a profit-maximizing environment. Detailed definitions of the considered variables, as well as detailed justification for the assumptions encoded in our DAG, are given in Tables 2 and 3 in the Appendix as well as in Section 2.

## 4.4 Identifiability considerations

The DAG shows the causal pathways through which CBI can affect consumer prices and thus ultimately inflation. We next explain the main paths from the intervention node to consumer prices. An independent central bank sets its policy tools autonomously to achieve its objective(s). Moreover, an independent central bank is less pressured to pursue an overly expansionary monetary policy that would produce only high inflation. Such a central bank is more likely to live up to its word, which increases its credibility (arrow 74).

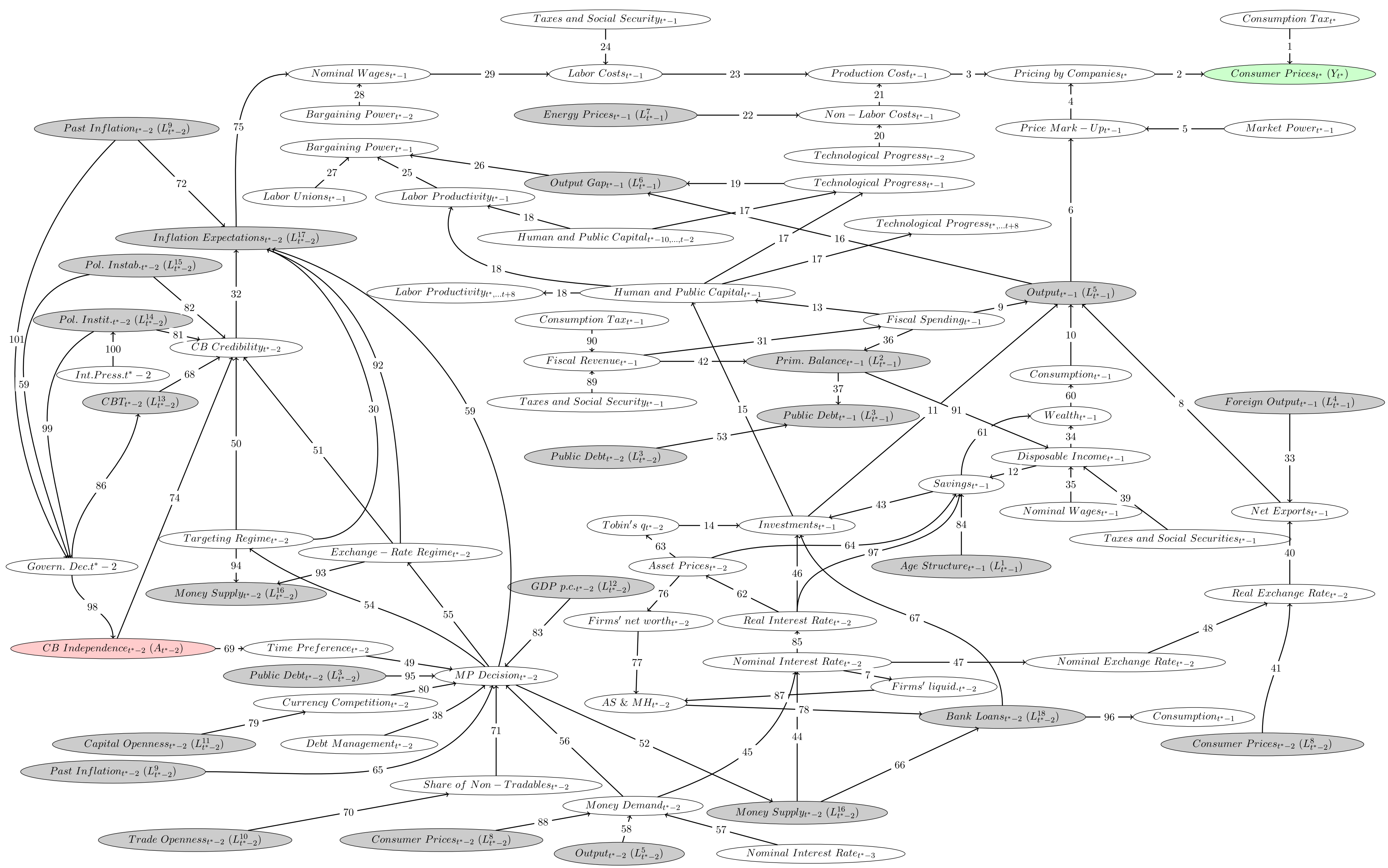


**Figure 1:** DAG containing the structural assumptions about the data generating process for a specific time point $t^* = 2000, \dots, 2010$. The target quantity is $\psi_{j,k}$ and relates to $Y_{2010}$, which refers to *Consumer Prices*$_{t^*}$ colored in green. The intervention rules relate to CBI at time $t^* - 2$, colored in red. Measured covariates are grey, and unmeasured covariates are white. A justification of the DAG is given in Appendix A.2.

Higher credibility keeps inflation expectations in check (arrow 32). The more contained inflation expectations are, the lower the demands for nominal wage compensation will be (arrow 75), which, in turn, keeps labor costs (arrow 29), production costs (arrow 23), and companies' prices (arrow 3) low. This will ultimately also be reflected in relatively low consumer prices (arrow 2). Another pathway from the intervention to the outcome acts through monetary policy decisions. Following an intervention, monetary policy makers' time preferences are reduced (arrow 69), and this will be taken into account in their monetary policy decisions (arrow 49). Monetary policy decisions are mirrored in money supply (arrow 52), which is tantamount to banks' loan creation (arrow 66) and, as a result, affects firms' investment decisions (arrow 67) and thus output (arrow 11). The final stage affects firms' markups (arrow 6) in their prices with a final effect on consumer prices (arrows 4 and 2).

There are several back-door paths from the intervention to the outcome. They all start with arrow 98 because CBI status is ultimately influenced by government decisions. Those decisions are affected by past inflation, political institutions and political stability (see Section 2 for a detailed justification). As an example, consider the back-door path that goes through government decisions (arrow 98) and past inflation (arrow 101): the latter affects current monetary policy decisions (arrow 65). Monetary policy will in turn impact the formation of inflation expectations (arrow 59) or the money supply (arrow 52). Along edges 66, 67, 11, 6, 4, and 2, this affects the outcome.

Under the assumption that the DAG as motivated in Appendix A is correct, establishing identification in terms of the (generalized) back-door criterion requires the following considerations: all back-door paths start with arrow 98 and can be blocked by conditioning on the following four variables: past inflation ($L_{t^*}^9$), central bank transparency ($L_{t^*}^{13}$), political institution ($L_{t^*}^{14}$), and political instability ($L_{t^*}^{15}$). There are various paths from the intervention to the outcome that start with edges 69, 49, and 52. All those paths contain mediators one should *not* necessarily condition on in our example because otherwise the effect of CBI on inflation through these paths would be blocked [40]. The same considerations apply to the paths starting with edges 74 and 32.

In summary, our DAG suggests that all back-door paths from $A_{t^*}$ to the outcome (that do not go through any future treatment node $A_{t^*+1}$) can be blocked by including $L_{t^*}^9$, $L_{t^*}^{13}$, $L_{t^*}^{14}$, and $L_{t^*}^{15}$ in the analysis. As many other variables lie on a mediating path from the intervention to the outcome (i.e., are descendants of $A_{t^*}$), they should not be conditioned upon.

We argue that the developed DAG should serve as the basis for identification considerations and estimation strategies. However, in complex macroeconomic situations, violations of this causal model need to be taken into account, and other estimation strategies may also be useful. We now explain how this can be facilitated.

## 4.5 Data-adaptive estimation with longitudinal TMLE

We can, in principle, follow the algorithm described in Section 3.7 to estimate the target quantity of interest. This includes estimation of the (nested) outcome model $\bar{Q}_{t^*}$ (step 1) and the intervention model $g_{0,A_{t^*}=\bar{d}_s^l}$ (step 3c) for each time point. That is, we can estimate the $g$-model for $t^* = 1998, \ldots, 2008$ and $Q_{t^*}$ for $t^* = 2000, \ldots, 2010$. As mentioned above, the DAG assumes a 2-year lag before an independent central bank can potentially affect the outcome. It is thus sufficient to estimate the first Q-model in 2000 given the assumed lag structure in the DAG. We define $Y_T := Y_{2010}$, which corresponds to the value of inflation in 2010, while $\bar{d}_{t^*}^1$, $\bar{d}_{t^*,i}^2(\bar{L}_{t^*-1}^8)$, and $\bar{d}_{t^*}^3$ are the interventions targeting CBI as described in Section 4.2.

We consider three approaches to covariate inclusion. The first is based on the identifiability considerations related to our DAG, and the other two refine variable inclusion criteria based on the scenario in which some structural causal assumptions in the DAG may be incorrect.

(i) **DAG-based approach** (*PlainDAG*): Based on the identifiability arguments from Section 4.4, $\bar{\mathbf{L}}_{t^*}$ contains only the relevant baseline variables from 1998 that were measured prior to the first intervention node, as well as $L_{t^*}^9$, $L_{t^*}^{13}$, $L_{t^*}^{14}$, and $L_{t^*}^{15}$.

(ii) **Greedy super learning approach** (*ScreenLearn*): This approach contains the full set of measured variables $\mathbf{L}_{t^*}$. This approach assumes that each variable could potentially lie on a back-door path but that this was undiscovered due to misspecification of the causal model. For example, a researcher who argues that bank loans directly affect a central bank's independence (i.e., that there is an arrow from bank loans to CBI) would have to consider a back-door path along arrows 67, 11, 6, 4, 2, and thus include public debt in $\mathbf{L}_{t^*}$. Similarly, if it is doubted that some variables are not necessarily mediators but rather confounders on a back-door path that exists due to unmeasured variables, e.g., *CBI* ← *unmeasured variable* → *Output* → … → *ConsumerPrices*, then measured variables such as *Output* (real GDP) would also have to be included in $\mathbf{L}_{t^*}$. We suggest that an analysis that includes all measured variables in $\mathbf{L}_{t^*}$ can serve as a useful sensitivity analysis to explore the extent to which effect estimates may change under different assumptions.

(iii) **Economic theory approach** (*EconDAG*): A further approach, termed *EconDAG*, includes only variables that are measured during a particular 2-yearly transmission cycle, as defined by our DAG. That is, for the Q-model at $t^*$, every measured variable between $t^* - 2$ and $t^* - 1$ is included, while for the estimation of the g-model at $t^* - 2$, only variables during the respective cycle are considered. As above, given the assumed time ordering, only variables from the past, and not from the future, are utilized in the respective models.

Given the complexity of the data-generating process, it makes sense to use machine learning techniques to estimate the respective g- and Q-models. For a specified set of learning algorithms and a given set of data, the method minimizing the expected prediction error (as estimated by $k$-fold cross validation) could be chosen. As the best algorithm in terms of prediction error may depend on the given data set, it is often recommended to use super learning instead – and this is what we use for (i), (ii), and (iii). Super learning [87] (or "stacking," [88]) considers a set of learners; instead of picking the learner with the smallest prediction error, one chooses the convex combination of learners that minimizes the $k$-fold cross validation error (for a given loss function, we use $k = 10$). The weights relating to this convex combination can be obtained with non-negative least squares estimation (which is implemented in the *R*-package `SuperLearner`, [89]). It can be shown that this weighted combination will perform asymptotically at least as well as the best algorithm, if not better, given that no correctly specified parametric model is contained in the set of learners [90].

As described in Section 3.7.2, the challenge of model specification, including the choice of appropriate learners and screening algorithms, is to address the complex nonlinear relationships in the data and the $p > n$ problem.

Our strategy is to use the following algorithms: the arithmetic mean of the outcome, generalized linear models (with main terms only and including all two-way interactions), Bayesian generalized linear models with an independent Gaussian prior distribution for the coefficients, classification and regression trees, multivariate adaptive (polynomial) regression splines, generalized additive models, Breimans' random forest, generalized boosted regression modeling, and single-hidden-layer neural networks. The algorithms are carefully chosen to reflect a balance between simple and computationally efficient strategies and more sophisticated approaches that are able to model highly nonlinear relationships and higher-order interactions that may be prevalent in the data. Furthermore, parametric, semiparametric, and nonparametric approaches were applied to allow for enough flexibility with respect to committing to parametric assumptions. In particular, tree-based procedures were chosen to handle challenges that frequently come with economic data – for instance outliers. In addition, since some of the continuous predictors are transformed by the natural logarithm, this strict monotone transformation may affect its variable importance in a regression-based procedure, while trees are not impaired in that respect.

For strategies (i)–(iii), we use the following learning and screening algorithms:

(a) **Screening algorithms**: Used only for estimation approach (ii) because of the large covariate set compared to the sample size; we used the elastic net [91], the random forest [92], Cramer's V (with either 4 or 8 variables selected at a maximum), and the Pearson correlation coefficient. The screening algorithms were chosen such that at least a subset of them could handle both categorical and quasi-continuous variables well.

(b) **Learning algorithms**: The 11 learning algorithms mentioned above are the same for estimation strategies (i) and (iii). (i) and (iii) were thus estimated with 11 algorithms each. In contrast, strategy (ii) additionally benefited from the five screening algorithms mentioned in (a) where each screening algorithms was run prior to each learning algorithm. We omitted generalized boosted regression modeling from the learner set such that $50 = 5 \times (11 - 1)$ algorithms (i.e., {Screener, Learner} tuples) emerged. In addition, learning algorithms that are applicable in the $p > n$ case were added without prior screening to the 50 tuples. As a result, when Breimans' random forest and single-hidden-layer neural networks were added without screening, 52 algorithms could be used for strategy (ii); see also Figure 4 in the Appendix.

All estimates have been obtained using the `ltmle` package in *R* [93].

## 4.6 Results

Descriptive summaries of the data are given in the Appendix, in Figures 6–8. They show the variables' distribution over time. Between 1998 and 2010 most measured variables show interesting patterns and changes. For example, one can see a continuously aging population in the countries included, as well as increased levels of central bank transparency. There is support in the data for all three treatment strategies, with 23 countries having an independent central bank throughout the whole time period, 27 countries never having an independent central bank for the period considered, and 16 countries which experienced periods with a negative median inflation rate or median inflation above 5% in the last 7 years during 1998 and 2010, while having legislated an independent central bank during the same time period (Figure 8).

A naive analysis comparing the mean reductions in inflation between 2000 and 2010 between those countries that had an independent central bank (from 1998 to 2008) and those that had a dependent central bank led to the following results: the mean reduction was 2.3 percentage points for those with an independent central bank, compared to 1.0 percentage points for those with a dependent central bank. This equates to a difference of 1.3 percentage points (95% CI: –6.1; 3.5). However, such a crude comparison does not allow a causal interpretation and is not an estimate of $\psi_{1,3}$.

The results of the analyses described in Section 4.5 are visualized in Figure 2.

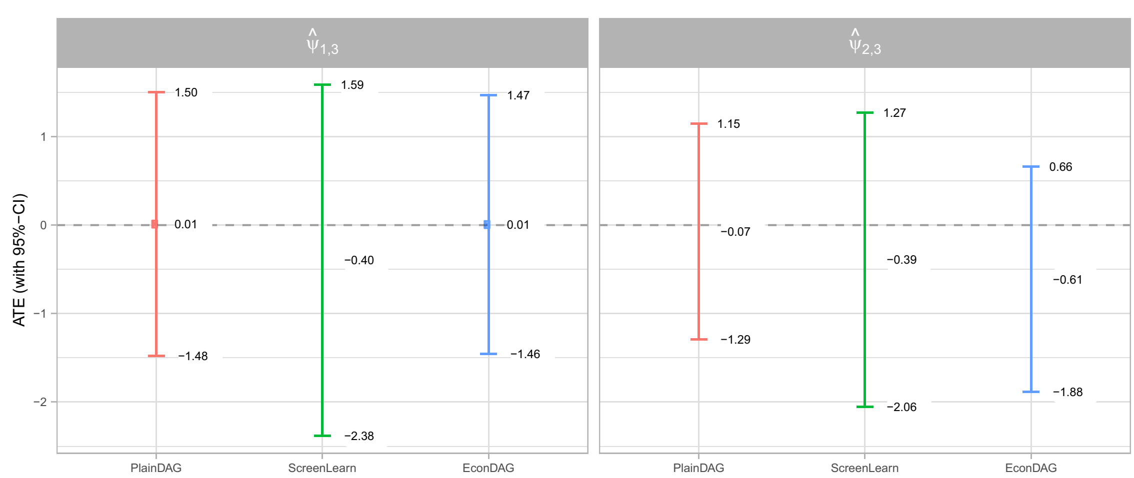


**Figure 2:** $\hat{\psi}_{1,3}$ and $\hat{\psi}_{2,3}$ for the three different estimation strategies.

Our main analysis (*PlainDAG*) suggests that if a country had legislated CBI for every year between 1998 and 2008, it would have had an average increase in inflation of 0.01 (95% confidence interval [CI]: –1.48; 1.50) percentage points in 2010. The other two approaches led to slightly different results: –0.44 (95% CI: –2.38; 1.59) for *ScreenLearn* and 0.01 (95% CI: –1.46; 1.47) for *EconDAG*.

Similarly, when considering the estimation strategy *PlainDAG*, we can conclude that if a country had legislated an independent central bank for every year when the median of the past 7 years of inflation had been above 5% or below 0% from 1998 to 2008, it would have achieved an average reduction in inflation of 0.07 percentage points (95% CI: –1.29; 1.15) in 2010 compared to a central bank that was independent during the same time span (i.e., dichotomized CBI = 0). The other two strategies suggest somewhat stronger inflation reductions.

Our findings can be summarized as follows: First, depending on the degree of structural assumptions imposed, we find that an independent central bank has either a negative or no effect on inflation. Second, as suggested by the confidence intervals, we cannot exclude the possibility of a strong negative or positive ATE. Third, the largest estimated ATE (in absolute terms) amounts to –0.61 percentage points (EconDAG). From a monetary policy perspective, this can be considered as substantial, given that our study period covers an era of overall low to moderate inflation (characterized by a median inflation rate of about 4%).

For a sensitivity analysis, we stratified our sample according to the World Bank's income classification into high income ($n = 26$) and low income ($n = 34$) countries and reran all analyses. The results are reported in the Appendix (cf. Figures 9 and 10). For high-income countries, the ATE (averaged across estimation strategies) is slightly positive. In contrast, for low-income countries, where inflation has typically been higher, we obtain almost no effect for the static treatment strategy (i.e., $\hat{\Psi}_{1,3}$) and an average reduction of about –0.4 percentage points for the dynamic treatment strategy $\hat{\Psi}_{2,3}$. However, due to the small sample sizes, these results need to be interpreted with caution.

The diagnostics for all analyses are given in Table 1 and Figure 5 in the Appendix. The cumulative product of inverse probabilities was never below the truncation level of 0.01, which was re-assuring [94]. The maximum value of clever covariates, as defined in (3), was always well below 5, which suggests that the

**Table 1:** Row 1: mean percentage of observations that had to be truncated because the cumulative product of inverse probabilities was <0.01. Rows 2 and 3: Mean and maximum value of the clever covariate. All results are averaged over the five imputed data sets. Rows 4 and 5 contain the minimum and maximum of the five mean clever covariate values across the imputed data sets

| Intervention | ScreenLearn, $\hat{\psi}_{1,3}$ | | ScreenLearn, $\hat{\psi}_{2,3}$ | | EconDAG, $\hat{\psi}_{1,3}$ | |
|---|---|---|---|---|---|---|
| | $\bar{A}_{t^*} = \bar{d}^3_{t^*}$ | $\bar{A}_{t^*} = \bar{d}^1_{t^*}$ | $\bar{A}_{t^*} = \bar{d}^3_{t^*}$ | $\bar{A}_{t^*} = \bar{d}^2_{t^*}$ | $\bar{A}_{t^*} = \bar{d}^3_{t^*}$ | $\bar{A}_{t^*} = \bar{d}^1_{t^*}$ |
| Trunc. (%) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| CC Mean | 0.89 | 0.90 | 0.91 | 0.52 | 0.83 | 0.71 |
| CC Max. | 3.64 | 4.88 | 3.71 | 2.57 | 2.27 | 2.63 |
| CC Mean Max. | 0.94 | 1.02 | 0.95 | 0.59 | 0.85 | 0.77 |
| CC Mean Min. | 0.81 | 0.77 | 0.88 | 0.48 | 0.79 | 0.62 |
| | EconDAG, $\hat{\psi}_{2,3}$ | | PlainDAG, $\hat{\psi}_{1,3}$ | | PlainDAG, $\hat{\psi}_{2,3}$ | |
| | $\bar{A}_{t^*} = \bar{d}^3_{t^*}$ | $\bar{A}_{t^*} = \bar{d}^2_{t^*}$ | $\bar{A}_{t^*} = \bar{d}^3_{t^*}$ | $\bar{A}_{t^*} = \bar{d}^1_{t^*}$ | $\bar{A}_{t^*} = \bar{d}^3_{t^*}$ | $\bar{A}_{t^*} = \bar{d}^2_{t^*}$ |
| Trunc. (%) | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| CC Mean | 0.83 | 0.50 | 0.82 | 0.72 | 0.82 | 0.50 |
| CC Max. | 2.30 | 1.97 | 2.25 | 2.63 | 2.26 | 1.99 |
| CC Mean Max. | 0.86 | 0.51 | 0.85 | 0.75 | 0.86 | 0.51 |
| CC Mean Min. | 0.79 | 0.47 | 0.78 | 0.64 | 0.79 | 0.47 |

chosen super learning approach worked well. However, the mean clever covariate, which is supposed to be broadly approximately 1, was not ideal for dynamic treatment strategy 2, suggesting that $\psi_{2,3}$ should be interpreted cautiously.

Figure 4 (Appendix) visualizes the learner weight distribution. In our analysis, a multitude of learners and screening algorithms were important, including neural networks, random forests, regression trees, and Bayesian generalized linear models.

# 5 Simulations

Motivated by our data analysis, we explore the extent to which model misspecification and choice of learner sets may affect effect estimation with longitudinal maximum likelihood estimation (and competing methods).

## 5.1 Data-generating processes

We specified two data-generating processes: a simple one with three time points and one time-dependent confounder and a more complex one with up to six time points and ten time-varying variables.

For the first simulation (*Simulation 1*), we assume the following time ordering:

$$O = (L_1, A_1, Y_1, L_2, A_2, Y_2, L_3, A_3, Y_3).$$

Using the $R$-package `simcausal` [95], we define preintervention distributions as listed in Table 4 (Appendix).

For the second simulation (*Simulation 2*), we use the following time ordering:

$$O = (L_1^1, A_1, Y_1, L_1^2, \ldots, L_1^{10}, \ldots, L_5^1, A_5, Y_5, L_5^2, \ldots, L_5^{10}, L_6^1, A_6, Y_6).$$

We generated the preintervention data according to the distributions specified in Table 5 (Appendix).

## 5.2 Target parameter and interventions

For both simulations, we were interested in evaluating ATEs between two static interventions. That is, we were interested in

$$\bar{d}_{t^+}^{Sim1,1} = \{a_{t^+} = 1 \quad \forall t^+ \in \{1, 2, 3\}$$
$$\bar{d}_{t^+}^{Sim1,0} = \{a_{t^+} = 0 \quad \forall t^+ \in \{1, 2, 3\}$$

and

$$\bar{d}_{t^{++}}^{Sim2,1} = \{a_{t^{++}} = 1 \quad \forall t^{++} \in \{1, 2, 3, 4, 5, 6\}$$
$$\bar{d}_{t^{++}}^{Sim2,0} = \{a_{t^{++}} = 0 \quad \forall t^{++} \in \{1, 2, 3, 4, 5, 6\}.$$

The target parameters of interest are thus

$$\psi_1 = \mathbb{E}\left(Y_2^{\bar{d}_{t^+}^{Sim1,1}}\right) - \mathbb{E}\left(Y_2^{\bar{d}_{t^+}^{Sim1,0}}\right), \quad \psi_2 = \mathbb{E}\left(Y_6^{\bar{d}_{t^{++}}^{Sim2,1}}\right) - \mathbb{E}\left(Y_6^{\bar{d}_{t^{++}}^{Sim2,0}}\right). \tag{8}$$

## 5.3 Estimations

In our primary analysis, we used LTMLE for both simulations. In a secondary analysis, we also evaluated the performance of (longitudinal) inverse probability of treatment weighting (see, e.g., Daniel et al. [18] and references therein).

For LTMLE, we considered four different estimation approaches, the first for the first simulation and another three for the second simulation:

(i) Estimation as explained in Section 3.7. Q- and g-models were fitted with (generalized) linear models. This is estimation approach *Generalized Linear Model* (GLM).

(ii) Estimation as explained in Section 3.7. Q- and g-models were fitted with a data-adaptive approach using super learning. There were four candidate learners: the arithmetic mean, GLMs, Bayesian generalized linear models with an independent Gaussian prior distribution for the coefficients, as well as classification and regression trees. No screening of variables was conducted. This is estimation approach *L1*.

(iii) Estimation as explained in Section 3.7. Q- and g-models were fitted with a data-adaptive approach using super learning. The same four learners as in *L1* are utilized; however, variable screening with Pearson's correlation coefficient was conducted. In addition, four more learners were added: multivariate adaptive (polynomial) regression splines [96], generalized additive models, and generalized linear models including the main effects with all corresponding two-way interactions. These additional four learners included variable screening with the elastic net ($\alpha = 0.75$). This is estimation approach *L2*.

(iv) Estimation as explained in Section 3.7. Q- and g-models were fitted with a data-adaptive approach using super learning. The eight learning/screening combinations from *L2* were used. In addition, single-hidden-layer neural networks were used, once without variable screening and once with elastic net screening. Finally, the last learner is composed of classification and regression with the random forest. This is estimation approach *L3*.

We also obtained estimates for the ATE based on IPTW. The estimation of the propensity scores was identical to the estimation of the g-models within LTMLE and is thus also based on the estimation procedures described in (i)–(iv).

## 5.4 Comparisons

We compared the estimated absolute (abs.) bias and coverage probabilities for the estimated ATEs for the two simulations and for both correctly and incorrectly specified Q-models (see details below).

(i) **Simulation 1**: The incorrect, misspecified, Q-models omit $\mathbf{L} := (L_1, L_2, L_3)$ entirely. By contrast, the g-models were specified such that the entire covariate histories are taken into account. As a result, if no screening is applied (estimation strategies GLM and L1), all relevant variables are used for estimation; however, with screening (estimation strategies L2 and L3), some variables might be omitted.

(ii) **Simulation 2**: The incorrect, misspecified, Q-models do not use $\mathbf{L}^1 := (L_1^1, L_2^1, L_3^1, L_4^1, L_5^1, L_6^1, L_7^1)$ for estimation. Thus, one relevant back-door path remains unblocked, which leads to time-dependent confounding with treatment-confounder feedback. As in simulation 1, all g-models were specified such that the entire covariate histories are taken into account.

## 5.5 Results

The results after 1,000 simulation runs are summarized in Figure 3.

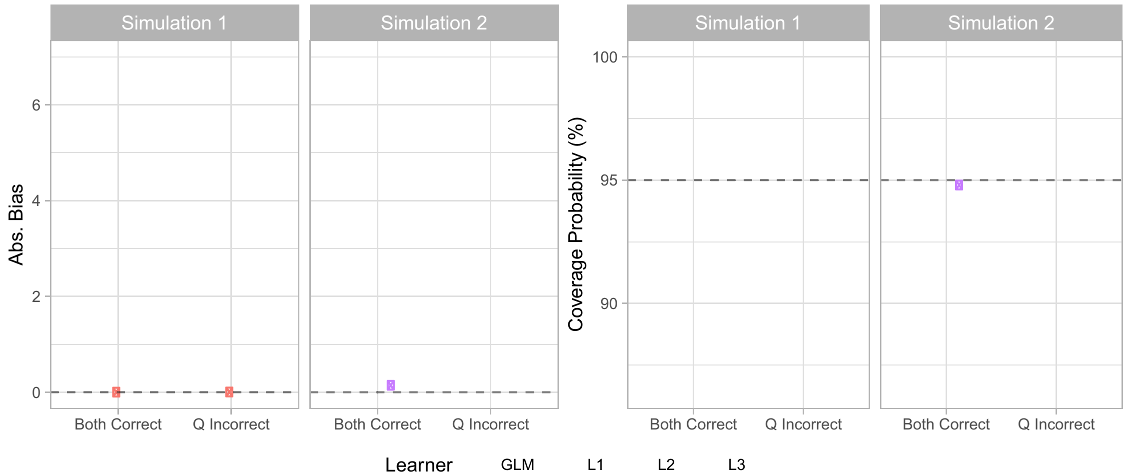


**Figure 3:** Absolute bias and coverage probability for both simulations – for correctly specified Q- and g-models (*Both Correct*) and misspecified Q-models (*Q Incorrect*) of LTMLE.

In simulation 1, LTMLE provides approximately unbiased estimates even under misspecified Q-models. This is because TMLE is a doubly robust estimator, and thus misspecification of either the Q- or g-models can be handled. However, the coverage probabilities are too high. See [74] for a discussion of this issue.

Under the more complex setup of simulation 2, there is small bias if both the Q- and g-models contain the relevant adjustment variables (*Both Correct*) and learner set *L1* is used (Bias = 0.991). The more sophisticated learner sets *L2* and *L3* yield much better estimates (Bias = 0.158 and 0.144). With incorrect specification of the Q-model, there is again some bias (Bias = 1.438, 0.639, 0.663). Interestingly, for simulation 2, the most complex estimation approach with the largest learner set *L3* does not produce a substantial improvement over *L2*. This highlights that a simple increase in learners does not necessarily improve the finite sample performance of LTMLE, although sufficient breadth and complexity are certainly always needed, as seen by the inferior performance of the first learner set.

In simulation 1, the confidence intervals have too large coverage probabilities. However, in simulation 2, using *L2* and *L3* yields (close to) nominal coverage probabilities. Nevertheless, our results highlight the need to develop more reliable variance estimators, such that overall better coverage can be achieved.

Note that while LTMLE may produce approximately unbiased point estimates, IPTW does not seem to benefit from complex estimation procedures for the propensity scores (g-models) in the second simulation. The estimates are rather volatile, with some bias and poor coverage probabilities. These conclusions hold for all learner sets considered (Appendix, Figure 11).

# 6 Conclusions

We have shown that even for complex macroeconomic questions, it is possible to develop a causal model and implement modern doubly robust longitudinal effect estimators. We believe that this is an important contribution in light of the current debate on the appropriate implementation and use of causal inference for economic questions [39]. Our suggestion was to commit to a causal model, motivate it in substantial detail (as in Appendix A.2), discuss possible violations of it, and ultimately conduct sensitivity analyses that evaluate effect estimates under different (structural) assumptions.

While the statistical literature has emphasized the benefits of doubly robust effect estimation in conjunction with extensive machine learning [22], its use in sophisticated longitudinal settings has sometimes been limited due to computational challenges and constraints [25]. We have shown how the use of screening and learning algorithms that are tailored to the question of interest can help to facilitate a successful implementation of this approach.

As stressed by Imbens [39]: "[...] *models in econometric papers are often developed with the idea that they are useful on settings beyond the specific application in the paper*". We hope that both our causal model, i.e., the DAG, and our proposed estimation techniques will be useful in applications other than ours.

Our simulation studies suggest that LTMLE with super learning can yield good point estimates compared to competing approaches, even under model misspecification. However, both the coverage of confidence intervals and the appropriate choice of learners are challenges that warrant more investigation. Recent research confirms that the development of more robust variance estimators is urgently needed [74] and that learner selection is becoming more diverse [97].

From a monetary policy point of view, we conclude that based on the estimates from the main analysis there is no strong support for the hypothesis that an independent central bank necessarily affects inflation, although our confidence intervals were wide. Making fewer or different structural assumptions, as in our secondary analyses, leads to an average inflation reduction of up to 0.6 percentage points under CBI. An ATE of −0.6 percentage points may be seen as substantial, considering that the period on which our estimations are based is overall characterized by low to moderate inflation. However, a naive use of super learning (as in our "ScreenLearn" secondary analysis) may be potentially dangerous because important collider and mediator structures may be overlooked, which can yield different, possibly incorrect results. A comparison of the point estimates from the main and secondary analysis reflects this consideration. As highlighted throughout this article, while a sophisticated computational approach can be advantageous for doubly robust causal effect estimation, it can not replace the commitment to well-thought-out structural assumptions about the macroeconomic process under consideration.

**Acknowledgement:** We acknowledge support from the University of Cape Town's Open Access Fund for this publication.

**Conflict of interest:** Authors state no conflict of interest.

**Data availability statement:** The data and code from this study are openly available at https://github.com/PFMB/CausalInflation. The use of the materials is permitted under the condition that attribution is given to the authors of this article.

## References

[1] Vuletin G, Zhu L. Replacing a "disobedient" central bank governor with a "docile" one: a novel measure of central bank independence and its effect on inflation. J Money Credit Banking. 2011;43(6):1185–215.
[2] Grilli V, Masciandaro D, Tabellini G. Political and monetary institutions and public financial policies in the industrial countries. Economic Policy. 1991;6(13):341–92.
[3] Cukierman A, Webb SB, Neyapti B. Measuring the independence of central banks and its effect on policy outcomes. World Bank Econ Rev. 1992;6(3):353–98.
[4] Alesina A, Summers LH. Central bank independence and macroeconomic performance: some comparative evidence. J Money Credit Banking. 1993;25(2):151–62.
[5] Klomp J, De Haan J. Inflation and central bank independence: a meta-regression analysis. J Econ Surv. 2010a;24(4):593–621.
[6] Klomp J, De Haan J. Central bank independence and inflation revisited. Public Choice. 2010b;144(3–4):445–57.
[7] Arnone M, Romelli D. Dynamic central bank independence indices and inflation rate: A new empirical exploration. J Financial Stabil. 2013;9(3):385–98.
[8] Cargill TF. The statistical association between central bank independence and inflation. BNL Quarter Rev. 1995;48(193):159–72.
[9] Fuhrer JC. Central bank independence and inflation targeting: monetary policy paradigms for the next millennium? New England Econ Rev. 1997;(Jan/Feb):19–36.
[10] Oatley T. Central bank independence and inflation: Corporatism, partisanship, and alternative indices of central bank independence. Public Choice. 1999;98(3–4):399–413.

[11] Neyapti B. Monetary institutions and inflation performance: cross-country evidence. J Econ Policy Reform. 2012;15(4):339–54.
[12] Alpanda S, Honig A. The impact of central bank independence on the performance of inflation targeting regimes. J Int Money Finance. 2014;44:118–35.
[13] Crowe C, Meade EE. The evolution of central bank governance around the world. J Econ Perspectives. 2007;21(4):69–90.
[14] Crowe C, Meade EE. Central bank independence and transparency: Evolution and effectiveness. Eur J Political Econ. 2008;24(4):763–77.
[15] Walsh CE. Optimal transparency under flexible inflation targeting. manuscript. Santa Cruz: University of California; 2005. URL https://iecon.tau.ac.il/sites/economy.tau.ac.il/files/media_server/Economics/Sapir/conferences/Carl%20E.Walsh.pdf.
[16] Jácome LI, Vázquez F. Is there any link between legal central bank independence and inflation? Evidence from Latin America and the Caribbean. Eur J Political Econ. 2008;24(4):788–801.
[17] Robins J. A new approach to causal inference in mortality studies with a sustained exposure period – application to control of the healthy worker survivor effect. Math Model. 1986;7(9–12):1393–512.
[18] Daniel RM, Cousens SN, De Stavola BL, Kenward MG, Sterne JA. Methods for dealing with time-dependent confounding. Stat Med. 2013;32(9):1584–618.
[19] Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000;11(5):550–60.
[20] Daniel RM, De Stavola BL, Cousens SN. gformula: Estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula. Stata Journal. 2011;11(4):479–517.
[21] Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. Biometrics. 2005;64(2):962–72.
[22] Van der Laan M, Rose S. Targeted Learning. New York: Springer; 2011.
[23] van der Laan MJ, Gruber S. Targeted minimum loss based estimation of causal effects of multiple time point interventions. Int J Biostat. 2012;8(1):9.
[24] Tran L, Yiannoutsos C, Wools-Kaloustian K, Siika A, van der Laan M, Petersen M. Double robust efficient estimators of longitudinal treatment effects: Comparative performance in simulations and a case study. Int J Biostat. 2019;15(2):20170054. doi: 10.1515/ijb-2017-0054.
[25] Schomaker M, Luque-Fernandez MA, Leroy V, Davies M-A. Using longitudinal targeted maximum likelihood estimation in complex settings with dynamic interventions. Stat Med. 2019;38(24):4888–911.
[26] Tinbergen J. Determination and interpretation of supply curves: an example. Zeitschrift für Nationalökonomie. 1930;1:669–79.
[27] Wright P. The method of path coefficients. Annals Math Stat. 1934;5:161–215.
[28] Hahn J, Todd P, Van der Klaauw W. Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica. 2001;69(1):201–9.
[29] Imbens G. Instrumental variables: An econometrician's perspective. Stat Sci. 2014;29(3):323–58.
[30] Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):688–701.
[31] Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, et al. Double/debiased machine learning for treatment and structural parameters. Econometric J. 2018;21(1):C1–C68. doi: 10.1111/ectj.12097.
[32] Kreif N, Tran L, Grieve R, deStavola B, Tasker R, Petersen M. Estimating the comparative effectiveness of feeding interventions in the paediatric intensive care unit: a demonstration of longitudinal targeted maximum likelihood estimation. Am J Epidemiol. 2017;186:1370–9.
[33] Decker A, Hubbard A, Crespi C, Seto E, Wang M. Semiparametric estimation of the impacts of longitudinal interventions on adolescent obesity using targeted maximum-likelihood: Accessible estimation with the ltmle package. J Causal Infer. 2014;2(1):95–108.
[34] Schnitzer ME, Moodie EE, van der Laan MJ, Platt RW, Klein MB. Modeling the impact of hepatitis C viral clearance on end-stage liver disease in an HIV co-infected cohort with targeted maximum likelihood estimation. Biometrics. 2014a;70(1):144–52.
[35] Schnitzer ME, van der Laan MJ, Moodie EE, Platt RW. Effect of breastfeeding on gastrointestinal infection in infants: a targeted maximum likelihood approach for clustered longitudinal data. Annals Appl Stat. 2014b;8(2):703–25.
[36] Schnitzer ME, Lok J, Bosch RJ. Double robust and efficient estimation of a prognostic model for events in the presence of dependent censoring. Biostat. 2016;17(1):165–77.
[37] Tran L, Yiannoutsos C, Musick B, Wools-Kaloustian K, Siika A, Kimaiyo S, et al. Evaluating the impact of a HIV low-risk express care task-shifting program: A case study of the targeted learning roadmap. Epidemiol Meth. 2016;5(1):69–91.
[38] Bell-Gorrod H, Fox MP, Boulle A, Prozesky H, Wood R, Tanser F, et al. The impact of delayed switch to second-line antiretroviral therapy on mortality, depending on failure time definition and cd4 count at failure. bioRxiv. 2019. doi: 10.1101/661629, URL https://www.biorxiv.org/content/biorxiv/early/2019/06/07/661629.full.pdf.
[39] Imbens G. Potential outcome and directed acyclic graph approaches to causality: relevance for empirical practice in economics. Working Paper 26104, National Bureau of Economic Research, MA 02138, U.S.A. July 2019; URL http://www.nber.org/papers/w26104.

[40] Hernan MA, Robins JM. Causal Inference, volume forthcoming of Chapman & Hall/CRC Monographs on Statistics & Applied Probab. Taylor & Francis; 2020. ISBN 9781420076165. URL https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/.
[41] Rogoff K. The optimal degree of commitment to an intermediate monetary target. Quarter J Econ. 1985;100(4):1169–89.
[42] Bernhard W, Lawrence Broz J, Clark WR. The political economy of monetary institutions. Int Org. 2002;56(4):693–723.
[43] Kern A, Reinsberg B, Rau-Göhring M. IMF conditionality and central bank independence. Eur J Political Econ. 2009;59(C):212–29.
[44] CarolinaGarriga A. Central bank independence in the world: A new data set. Int Interact. 2016;42(5):849–68.
[45] Dincer NN, Eichengreen B. Central bank transparency and independence: Updates and new measures. Int J Central Banking. 2014;10(1):189–259.
[46] Baumann P, Rossi E, Volkmann A. What drives inflation and how? Evidence from additive mixed models selected by cAIC. Working Paper, 2019. URL https://arxiv.org/pdf/2006.06274.pdf.
[47] Farvaque E. Political determinants of central bank independence. Econ Lett. 2002;77(1):131–5.
[48] Aghion P, Alesina A, Trebbi F. Endogenous political institutions. Quarter J Econ. 2004;119(2):565–611.
[49] Polillo S, Guillén MF. Globalization pressures and the state: The worldwide spread of central bank independence. Am J Sociol. 2005;110(6):1764–802.
[50] Brumm HJ. The effect of central bank independence on inflation in developing countries. Econ Lett. 2006;90(2):189–93.
[51] Brumm HJ. Inflation and central bank independence: Two-way causality? Econ Lett. 2011;111(3):220–2.
[52] Bodea C, Hicks R. Price stability and central bank independence: Discipline, credibility, and democratic institutions. Int Org. 2015;69(1):35–61.
[53] Romelli D. The political economy of reforms in central bank design: evidence from a new dataset. BAFFI CAREFIN Centre Research Paper. 2018;87.
[54] Masciandaro D, Romelli D. Peaks and troughs: economics and political economy of central bank independence cycles. The Oxford Handbook of the Economics of Central Banking; New York, NY, United States of America: Oxford University Press; 2019.
[55] de Hann J, Jan van't Hag G. Variation in central bank independence across countries: some provision empirical evidence. Public Choice. 1995;85(3–4):335–51. URL https://ideas.repec.org/a/kap/pubcho/v85y1995i3-4p335-51.html
[56] Moser P. Checks and balances, and the supply of central bank independence. Eur Econ Rev. 1999;43(8):1569–93.
[57] Cukierman A, Webb SB. Political influence on the central bank: international evidence. World Bank Econ Rev. 1995;9(3):397–423.
[58] Wachtel P, Blejer MI. A fresh look at central bank independence. Cato J. 2020;40:105.
[59] Blejer MI, Leone AM, Rabanal P, Schwartz G. Inflation targeting in the context of imf-supported adjustment programs. IMF Staff Papers. 2002;49(3):313–38.
[60] Gutiérrez E. Inflation performance and constitutional central bank independence: evidence from Latin America and the caribbean. IMF Working Paper 03/53, 2003.
[61] Rodrik D. Goodbye Washington consensus, hello Washington confusion? a review of the world bank's economic growth in the 1990s: learning from a decade of reform. J Econ Liter. 2006;44(4):973–87.
[62] Reinsberg B, Stubbs T, Kentikelenis A, King L. Bad governance: How privatization increases corruption in the developing world. Regul Govern. 2020;14(4):698–717.
[63] Simmons BA, Elkins Z. The globalization of liberalization: Policy diffusion in the international political economy. Am Polit Sci Rev. 2004;98(1);171–89.
[64] Elhorst P, Zandberg E, De Haan J. The impact of interaction effects among neighbouring countries on financial liberalization and reform: A dynamic spatial panel data approach. Spatial Econ Anal. 2013;8(3):293–313.
[65] Giuliano P, Mishra P, Spilimbergo A. Democracy and reforms: evidence from a new dataset. Am Econ J Macro Econ. 2013;5(4):179–204.
[66] Acemoglu D, Naidu S, Restrepo P, Robinson JA. Democracy does cause growth. J Polit Econ. 2019;127(1):47–100.
[67] Daniel RM, De Stavola BL, Cousens SN. G-formula: Estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula. Stata J. 2011;11(4):479–517.
[68] Robins J, Hernan MA. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G, editors. Longitudinal data analysis. CRC Press; 2009. p. 553–99.
[69] Young JG, Cain LE, Robins JM, O'Reilly EJ, Hernan MA. Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula. Stat Biosci. 2011;3(1):119–43.
[70] Pearl J. An introduction to causal inference. Int J Biostat. 2010;6(2):7.
[71] Molina J, Pantazis L, Sued M. Some considerations on the back door theorem and conditional randomization. Epidemiol Meth. 2014;3(1):113–20.
[72] Richardson T, Robins J. Single world intervention graphs (SWIGs): a unification of the counterfactual and graphical approaches to causality. Center for Statistics and the Social Sciences, University of Washington Working Paper Series. Number 128, 2013. URL https://www.csss.washington.edu/Papers/wp128.pdf.
[73] Schnitzer ME, Cefalu M. Collaborative targeted learning using regression shrinkage. Stat Med. 2017;37:530–43.
[74] Tran L, Petersen M, Schwab J, van der Laan MJ. Robust variance estimation and inference for causal effect estimation. arXiv e-prints, arXiv:1810.03030. Oct 2018. URL https://arxiv.org/abs/1810.03030.

[75] van der Laan MJ, Rose S. Targeted learning in data science: causal inference for complex longitudinal studies. Springer Series in Statistics. Springer International Publishing; 2018. ISBN 9783319653044.
[76] Schnitzer ME, Lok J, Gruber S. Variable selection for confounder control, flexible modeling and collaborative targeted minimum loss-based estimation in causal inference. Int J Biostat. 2016;12(1):97–115.
[77] Pearl J. Invited commentary: understanding bias amplification. Am J Epidemiol. 174(11):1223–7; discussion p. 1228–9, 2011. ISSN 1476-6256 (Electronic) 0002-9262 (Linking). doi: 10.1093/aje/kwr352 URL https://www.ncbi.nlm.nih.gov/pubmed/22034488.
[78] Honaker J, King G, Blackwell M. Amelia II: a program for missing data. J Stat Software. 2011;45(7):1–47.
[79] Friedman M. Have monetary policies failed? Am Econ Rev. 1972;62(1/2):11–8.
[80] Batini N, Nelson E. The lag from monetary policy actions to inflation: Friedman revisited. Int Finance. 2001;4(3):381–400.
[81] Goodhart CAE. Monetary transmission lags and the formulation of the policy decision on interest rates. In: Challenges for Central Banking. Springer; 2001. 205–28.
[82] Taylor JB, Wieland V. Surprising comparative properties of monetary models: results from a new model database. Rev Econ Stat. 2012;94(3):800–16.
[83] Havranek T, Rusnak M. Transmission lags of monetary policy: a meta-analysis. Int J Cent Banking. 2013;9(4):39–76.
[84] Tobin J. Money and economic growth. Econometrica. 1965;33(4):671–84. ISSN 00129682, 14680262.
[85] Fisher I. The debt-deflation theory of great depressions. Econometrica: J Econ Soc. 1933;1(4):337–57.
[86] Mishkin FS. Rethinking monetary policy after the crisis. J Int Money Finance. 2017;73:252–74.
[87] Van der Laan MJ, Polley EC, Hubbard AE. Super learner. Stat Appl Genetics Mol Biol. 2007;6(1):1–21.
[88] Breiman L. Stacked regressions. Machine Learn. 1996;24(1):49–64.
[89] Polley E, LeDell E, Kennedy C, van der Laan M. SuperLearner: Super Learner Prediction. R package version 2.0-22. 2017.
[90] Van der Laan M, Polley E, Hubbard A. Super learner. Stat Appl Genet Mol Biol. 2008;6:25.
[91] Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc B (Statistical Methodology). 2005;67(2):301–20.
[92] Breiman L. Random forests. Machine Learn. Oct 20014;5(1):5–32.
[93] Lendle SD, Schwab J, Petersen ML, van der Laan MJ. ltmle: An R package implementing targeted minimum loss-based estimation for longitudinal data. J Stat Software. 2017;81(1):1–21. doi: 10.18637/jss.v081.i01.
[94] Petersen ML, Porter KE, Gruber S, Wang Y, van der Laan MJ. Diagnosing and responding to violations in the positivity assumption. Stat Meth Med Res. 2012;21(1):31–54.
[95] Sofrygin O, van der Laan MJ, Neugebauer R. Simcausal: simulating longitudinal data with causal inference applications. R package version 0.5.3. 2016.
[96] Friedman JH. Multivariate adaptive regression splines. Ann Stat. 1991;19(1):1–67.
[97] Gehringer C, Rode H, Schomaker M. The effect of electrical load shedding on pediatric hospital admissions in South Africa. Epidemiology. 2018;29(6):841–7.
[98] Gelardi A. Value added tax and inflation: a graphical and statistical analysis. Asian J Fin Account. 2014;6(1):138–58.
[99] Burda M, Wyplosz C. Macroeconomics: a European Text. Oxford: OUP; 2010.
[100] Bloch H, Olive M. Pricing over the cycle. Rev Indust Org. 2001;19(1):99–108.
[101] Mishkin FS, Matthews K, Giuliodori M. The Economics of Money, Banking and Financial Markets: European Edition. Pearson; 2013.
[102] Blanchard O, Amighini A, Giavazzi F. Macroeconomics: A European Perspective. Harlow, England: Prentice Hall; 2010.
[103] Diebolt C, Hippe R. The long-run impact of human capital on innovation and economic development in the regions of Europe. Appl Econ. 2019;51(5):542–63.
[104] Cahuc P, Postel-Vinay F, Robin J-M. Wage bargaining with on-the-job search: Theory and evidence. Econometrica. 2006;74(2):323–64.
[105] Gürkaynak RS, Levin A, Swanson E. Does inflation targeting anchor long-run inflation expectations? Evidence from the US, UK, and Sweden. J Europ Econ Ass. 2010;8(6):1208–42.
[106] Walsh CE. Monetary theory and policy. 3rd ed. Cambridge, Massachusetts: The MIT Press; 2010.
[107] Bernanke B, Laubach T, Mishkin F, Posen A. Inflation targeting: lessons from the international experience. Princeton, NJ: Princeton University Press; 2001.
[108] Heijdra BJ, van der Ploeg F. The foundations of modern macroeconomics. Oxford: Oxford University Press; 2002.
[109] Burnside C. Fiscal sustainability in theory and practice: a handbook. Washington, DC: World Bank; 2005.
[110] Missale A, Blanchard O. The debt burden and debt maturity. Am Econ Rev. 1994;84(1):309–19.
[111] OECD. 2020. Taxing Wages 2020. doi: 10.1787/047072cd-en, URL https://www.oecd-ilibrary.org/content/publication/047072cd-en.
[112] Eijffinger S, de Haan J. The political economy of central-bank independence. Princeton studies in international economics, International Economics Section, Departement of Economics Princeton University; 1996.
[113] Svensson LEO. Inflation targeting. In: Handbook of monetary economics. vol. 3, Elsevier; 2010. p. 1237–302.
[114] Mishkin FS. International experiences with different monetary policy regimes. J Monetary Econ. 1999;43(3):579–605.
[115] Cooper D, Dynan K. Wealth effects and macroeconomic dynamics. J Econ Surveys. 2016;30(1):34–55.
[116] Disney R, Gathergood J, Henley A. House price shocks, negative equity, and household consumption in the United Kingdom. J Euro Econ Ass. 2010;8(6):1179–207.

[117] Taylor JB. Discretion versus policy rules in practice. In: Carnegie-Rochester Conference Series on Public Policy. 1993. vol. 39. p. 195–214.
[118] Lane P. Inflation in open economies. J Int Econ. 1997;42(3–4):327–47.
[119] Romer D. Openness and inflation: theory and evidence. Quarter J Econ. 1993;108(4):869–903.
[120] Carlin W, Soskice D. Macroeconomics: Institutions, instability, and the financial system. Oxford, United Kingdom; New York, NY, United States of America: Oxford University Press; 2015.
[121] Blinder AS. Central-bank credibility: why do we care? how do we build it? Am Econ Rev. 2000;90(5):1421–31.
[122] Wei S-J, Tytell I. Does financial globalization induce better macroeconomic policies? IMF Working Paper. 2004;4(84). URL https://www.imf.org/en/Publications/WP/Issues/2016/12/30/Does-Financial-Globalization-Induce-Better-Macroeconomic-Policies-17267.
[123] Keefer P, Stasavage D. The limits of delegation: Veto players, central bank independence, and the credibility of monetary policy. Am Political Sci Rev. 2003;97(3):407–23. doi: 10.1017/S0003055403000777.
[124] Hayo B, Voigt S. Inflation, central bank independence, and the legal system. J Instit Theoret Econ (JITE). 2008;164(4):751–77.
[125] Campillo M, Miron JA. Why does inflation differ across countries? Working Paper 5540, National Bureau of Economic Research, MA 02138, U.S.A. April 1996. URL http://www.nber.org/chapters/c8889.pdf.
[126] Bobeica E, Nickel C, Lis E, Sun Y. Demographics and inflation. Technical report, ECB Working Paper. 2017. URL https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp2006.en.pdf?de1664eb9174babfb81c92cb90952fbc
[127] Elmendorf D, Gregory Mankiw N. Government debt. In: Taylor JB, Woodford M, editors. Handbook of macroeconomics. vol 1, Part C, chapter 25, 1st edition, Elsevier; 1999. p. 1615–69.
[128] Svensson LEO. Inflation forecast targeting: implementing and monitoring inflation targets. European Econ Rev. 1997;41(6):1111–46.
[129] Jordan T, Peytrignet M, Rossi E. Ten years experience with the Swiss National Bank monetary policy strategy. Swiss J Econ Stat. 2010;146(1):9–90.
[130] Kwon G, McFarlane L, Robinson W. Public debt, money supply, and inflation: a cross-country study. IMF Staff Papers. Aug 2009;56(3):476–515. ISSN 1564-5150. doi: 10.1057/imfsp.2008.26.
[131] Bacchetta P, Gerlach S. Consumption and credit constraints: International evidence. J Monet Econ. 1997;40(2):207–38. ISSN 0304-3932. doi: 10.1016/S0304-3932(97)00042-1.

# Appendix A More details on the causal model

## A.1 Definition of the variables listed in the DAG

**Table 2:** Explanation and definition of the variables shown in the causal model

| Node | Explanation | Emp. Approx. |
|---|---|---|
| Consumer Prices | Price changes (%) in the consumption basket of a representative household. | ConsumerPrices |
| Consumption Tax | Value added tax on the net price of goods and services. | Unmeasured |
| Pricing by Companies | Firms set their product prices based on production costs and markups to maximize profit. | Unmeasured |
| Price Mark-up | Surcharge on marginal cost. It depends on aggregate demand and market power. | Unmeasured |
| Production Cost | Convenient breakdown of unit costs into labor and nonlabor costs. It generally depends on the industry and countries' development. | Unmeasured |
| Labor Cost | Direct wages, salaries, labor taxes, and social security contributions. | Unmeasured |
| Nonlabor Cost | Capital, land, and intermediate inputs such as intermediate goods, primary commodities, and energy. | Unmeasured |
| Energy Prices | Weighted average of world market prices for energy resources such as oil, gas, and coal measured in USD. | EnergyPrices |
| Taxes and Social Security | Labor taxes and social security contributions. | Unmeasured |
| Market Power | Perfect competition forces firms to set marginal costs equal to prices. This corresponds to a lack of market power. By contrast, product differentiation suggests high market power. | Unmeasured |
| Output | Real GDP measured in USD. In a small open economy, output consists of consumption, investments, government spending, and net exports. | Output |
| Consumption | Private consumption as a share of disposable household income. This is divided into two components: autonomous consumption and marginal propensity to consume. | Unmeasured |
| Disposable Income | Consumer income after transfers and taxes. | Unmeasured |
| Tobin's *q* | An economic measure that compares the market value of installed capital with the replacement cost of installed capital. A value greater than 1 leads to new investments. If the value is smaller than 1, purchasing existing capital is cheaper than investing in new capital. | Unmeasured |
| Investments | Purchases of real estate by households and purchases of new capital goods (machines and plants) by firms. | Unmeasured |
| Nominal Wages | Employees' salaries unrelated to the development of prices or indexation. | Unmeasured |
| Bargaining Power | Strength of bargaining position of employees in the wage-setting process. | Unmeasured |
| Labor Unions | Associations that represent the employed labor force in setting wage levels, working conditions, and worker rights. | Unmeasured |
| Labor Productivity | The ratio of output (GDP) to the number of workers. | Unmeasured |
| Output Gap | Fluctuations (%) of current output (GDP) from its potential. | OutputGap |
| Technological Progress | A technological improvement resulting in higher machine productivity. | Unmeasured |
| Human and Public Capital | Expenses for discovering and developing new ideas and products. | Unmeasured |
| Inflation Expectations | Expected consumer price level changes (%) approximated by the backward-looking geometric mean of consumer price changes (%) over the past 3 years. | InflationExpectations |
| Savings | The sum of accumulated private (and public) savings. Savings can be negative. | Unmeasured |
| Foreign Output | World output (GDP) depending on foreign consumption, investment, and fiscal spending. Measured in USD as the sum of outputs for every country in the sample minus the output of the particular country. | ForeignOutput |
| Net Exports | Defined as exports minus the value of imports. | Unmeasured |
| Real Exchange Rate | Determined by the nominal exchange rate and the domestic and foreign price levels. | Unmeasured |

**Table 2:** continued

| Node | Explanation | Emp. Approx. |
|---|---|---|
| Nominal Exchange Rate | Domestic currency in terms of foreign currency. | Unmeasured |
| Fiscal spending | The sum of all government expenditures (on education, consumption, investments, etc.). | Unmeasured |
| Fiscal Revenue | The sum of fiscal earnings (mainly taxes). | Unmeasured |
| Primary Balance | Primary surplus/deficit: Government revenues minus government spending excluding interest payments on outstanding debt in percent of GDP. | PrimaryBalance |
| Public Debt | If the government runs a primary deficit in a given year, debt increases. The increase in debt is exacerbated by interest payments on existing debt. Measured in percent of GDP. | PublicDebt |
| Debt Management | Decisions of a government on debt structure, potentially resulting in different currency, price, and interest-rate indexation composition as well as different maturities of newly issued and outstanding debt. | Unmeasured |
| Money Demand | Demand for money, defined as currency plus deposit accounts, determined by GDP and the level of interest rates on bonds. | Unmeasured |
| Money Supply | Different monetary aggregates (M0–M3) are available. For this analysis M2 was used. Measured average annual growth rate (%) in money and quasi money. | MoneySupply |
| Nominal Interest Rate | The level of the interest rate is determined by the intersection of money supply and money demand. | Unmeasured |
| Targeting Regime | Monetary policy strategy introduced in the 1990s intended to stabilize inflation at a pre-announced point target or target range. | Unmeasured |
| Exchange-rate Regime | Monetary policy strategy intended to stabilize inflation at a level commensurate with that of a strong currency. By pegging the currency to an anchor country's currency, its monetary policy and, hence, inflation are imported. Deviations from the target exchange rate are corrected by purchases and sales of the pegged currency. | Unmeasured |
| Capital Openness | Index measuring a country's degree of capital account openness. | CapitalOpenness |
| AS and MH | Adverse selection and moral hazard due to information asymmetries in credit markets. | Unmeasured |
| Firms' Net Worth | A firm's total assets minus its total liabilities yields its equity. | Unmeasured |
| Firms' Liquidity | Firms' liquidity is directly linked to their cashflow. Cash is the most liquid asset and is used to meet short-term liabilities. | Unmeasured |
| Age Structure | Demographic indicator that captures the share of the total population older than 65 years. | AgeStructure |
| Trade Openness | The sum of imports and exports is set in relation to a country's output (%). It is a proxy for globalization. | TradeOpenness |
| Asset Prices | Prices of assets in which households, firms, or governments are able to hold wealth, such as stocks, bonds, bank deposits, cash or real estate. | Unmeasured |
| Real Interest Rate | The difference between the nominal interest rate and the expected rate of inflation. | Unmeasured |
| Currency Competition | Governments and central banks are forced to implement disciplined policies since they compete with foreign currencies for capital. The primary mechanism through which greater openness to foreign capital might lead to lower inflation arises presumably from its disciplining effect on monetary policy. | Unmeasured |
| CBT | Central banks publicly announce their forecasts, policy decisions, and assessments of the economy. A central bank's transparency is strongly related to its accountability and its credibility. Its measured by means of a numeric index ranging from 0 (least transparent) to 15 (most transparent). | CBTransparency |
| CBI | Independence of a central bank from governmental bodies. Measured via de jure indices (e.g., statutes); see the main text for detailed explanations. Dichotomized. | CBI |

**Table 2:** continued

| Node | Explanation | Emp. Approx. |
|---|---|---|
| CB Credibility | A central bank that does what it has announced publicly is considered to be credible. This is reflected in inflation expectations that are low and stable. | Unmeasured |
| Pol. Instab. | The percentage of veto players dropping from the government in any given year. In presidential systems, veto players are defined as the president and the largest party in the legislature. In parliamentary systems, the veto players are defined as the prime minister and the three largest government parties. | PolInstability |
| Pol. Instit. | A variety of dimensions: the stability orientation of society, the age of a country's political parties, average government duration, democratization and economic liberalization, the degree of civil liberties, the presence of a higher chamber and federalism as well as a legislative process characterized by extensive checks and balances. Measured as a ordered categorical variable with seven categories where "7" represents the lowest degree of civil liberties and "1" the highest. | PolInstitution |
| Time Preference | Time horizon envisaged by policymakers within which they want to achieve a certain macroeconomic outcome. It may vary from a short (high time preference) to a middle- to long-term perspective (low time preference). | Unmeasured |
| Share of Non-tradables | Distinction between tradeable and non-tradeable goods. Non-tradability means that a good is produced and consumed in the same economy (e.g., haircuts). | Unmeasured |
| GDP p.c. | GDP is the sum of all finished goods and services that are produced in a year. The p.c. term divides this value by the number of citizens. GDP p.c. is a proxy for economic wealth and living standards and is measured in USD. | GDPpc |
| Bank Loans | Commercial banks create money when they offer loans depending on the availability of central bank reserves at their disposal. Measured as domestic credit to private sector (% of GDP). | BankLoans |
| Past Inflation | Consumer price changes (%) during the past 7 years. "PastInflation" takes the median over those years. | PastInflation |
| MP Decision | Monetary policy makers' (i.e., central bankers') decisions are contractionary, neutral, or expansionary. | Unmeasured |
| Wealth | Household wealth is accumulated savings over previous periods (it can be negative in the event of net debt) and disposable income in the current period. | Unmeasured |
| Int. Pressure | International institutions or foreign aids can provide an important incentive to reform the central bank. For example, the IMF regularly demands countries to end monetary financing of public debt, remove central bank governors and board members, and sometimes even pushes for full-fledged central bank reform. Further pressure may be exerted by the (regional) peer CBI community and international investors. | Unmeasured |
| Govern. Dec. | Captures all decisions a government takes but which are limited to CBI. | Unmeasured |

## A.2 Explanation for the arrows in the DAG

**Table 3:** Explanation and references for the arrows shown in the causal model

| Arrow | Causality assumption | Source |
|---|---|---|
| 1 | Consumer prices can change after changes in consumption taxes (e.g., VAT). | Gelardi [98] |
| 2 | Consumer prices are set individually by retailers and companies. | Burda and Wyplosz [99, p. 290] |
| 3 | Production costs generally dominate the price-setting process. Profit margins strongly depend on the industry in question. | Burda and Wyplosz [99, p. 291] |
| 4 | Channels the aggregate demand side of the price-setting process. In a small open economy, demand shocks to goods and services affect the price level. | Burda and Wyplosz [99, p. 312] |
| 5 | Higher product differentiation leads to higher market power and higher markups in a profit-maximizing environment. | Burda and Wyplosz [99, p. 291] |
| 6 | Changes in aggregate demand in the goods market enable firms to set higher prices. | Bloch and Olive [100] |
| 7 | Expansionary monetary policy, which lowers nominal interest rates, also causes an improvement in firms' balance sheets because it raises their cash flow. The rise in cash flow increases firms' (or households') liquidity. | Mishkin et al. [101, p. 544 f.] |
| 8 | In a small open economy, domestic demand for goods, and thus output, is also affected by net exports. | Blanchard et al. [102, p. 125] |
| 9 | Fiscal spending describes the decision of the government to spend money. It affects output (GDP). | Blanchard et al. [102, p. 45] |
| 10 | Private consumption also affects output. | Blanchard et al. [102, p. 44] |
| 11 | Investments are another factor affecting output. | Blanchard et al. [102, p. 44] |
| 12 | The share of disposable income that is not consumed in this period is saved based on the marginal propensity to save. | Blanchard et al. [102, p. 52] |
| 13 | Governments undertake investments in human capital (e.g., education) or public capital (e.g., infrastructure) to bolster long-run economic growth. | Burda and Wyplosz [99, pp. 85 ff.] |
| 14 | A Tobin's $q$ not equal to 1 gives incentives to invest or divest in capital and therefore affects aggregate investment. | Burda and Wyplosz [99, p. 195] |
| 15 | Similar to arrow 13, companies and other non-governmental agents affect human capital. | Burda and Wyplosz [99, pp. 85 ff.] |
| 16 | The current value of GDP may deviate from its potential. | Burda and Wyplosz [99, p. 11] |
| 17 | Investments in human capital have a positive impact on innovation and economic development. | Diebolt and Hippe [103] |
| 18 | Training and education generally lead to high-skilled workers, and in turn, to high productivity of the labor force. | Burda and Wyplosz [99, pp. 85 f.] |
| 19 | Potential output growth is mainly determined by technological progress. | Burda and Wyplosz [99, p. 71] |
| 20 | Technological progress indicates higher productivity, and higher productivity can again be expressed as obtaining the same output with fewer inputs (here, lower nonlabor costs and higher profits) | Burda and Wyplosz [99, p. 71] |
| 21 | The first (second cf. 23) component that determines the production costs are nonlabor costs. | Burda and Wyplosz [99, p. 291] |
| 22 | Changes in energy prices are transmitted through supply shocks and affect the nonlabor costs of production. | Burda and Wyplosz [99, p. 297] |

**Table 3:** continued

| Arrow | Causality assumption | Source |
|---|---|---|
| 23 | The second component that determines production costs is labor costs. | Burda and Wyplosz [99, p. 291] |
| 24 | Gross hourly labor costs also include vacation, social security contributions, and other benefits paid by employers to the benefit of workers. | Burda and Wyplosz [99, p. 291] |
| 25 | Higher skills increase workers' bargaining power in the wage-setting process. | Cahuc et al. [104] |
| 26 | During boom periods, rising employment generally improves the bargaining position of workers. | Burda and Wyplosz [99, p. 294] |
| 27 | Labor unions generally improve the bargaining position of workers. | Burda and Wyplosz [99, p. 121] |
| 28 | A better bargaining position leads to higher wage markup. | Burda and Wyplosz [99, p. 294] |
| 29 | Nominal wages translate directly into labor costs. | Burda and Wyplosz [99, p. 292] |
| 30 | Inflation expectations are built on publicly announced inflation targets. | Gürkaynak et al. [105] |
| 31 | Fiscal revenue increases the government's capacity to spend. | Walsh [106, p. 136] |
| 32 | A credible central bank commitment to low and stable inflation anchors long-run inflation expectations. | Bernanke et al. [107] |
| 33 | Net exports depend positively on foreign output. | Blanchard et al. [102, p. 125] |
| 34 | Wealth depends on disposable income. | Heijdra and van der Ploeg [108, p. 136] |
| 35 | Disposable income is also determined by wages. | Blanchard et al. [102, p. 43] |
| 36 | Fiscal spending affects the primary balance. | Blanchard et al. [102, p. 439] |
| 37 | Current primary deficits are financed by new debt. | Burnside [109, p. 12] |
| 38 | Missale and Blanchard (1994) introduced what they called "effective maturity." Effective maturity measures the sensitivity of debt to unexpected inflation. The lower it is, the lower the impact of surprise inflation on the value of the debt, and the lower the incentive to inflate. | Missale and Blanchard [110] |
| 39 | Disposable income decreases following an increase in taxes and social security contributions and increases after a reduction in taxes and social security contributions. | OECD [111] |
| 40 | Exports depend negatively on the real exchange rate. | Blanchard et al. [102, p. 125] |
| 41 | The real exchange rate is partly determined by the domestic price level. | Blanchard et al. [102, p. 112] |
| 42 | Fiscal revenue affects the primary balance. | Blanchard et al. [102, p. 439] |
| 43 | Investments are proportional to output. Higher output implies higher savings and, thus, higher investments. | Blanchard et al. [102, p. 248] |
| 44 | The interest rate is determined by the equilibrium condition that the supply of money be equal to the demand for money. | Blanchard et al. [102, p. 77] |
| 45 | The interest rate is determined by the equilibrium condition that the supply of money be equal to the demand for money. | Blanchard et al. [102, p. 77] |
| 46 | "An important feature of the interest-rate transmission mechanism is its emphasis on the real (rather than the nominal) interest rate as the rate that affects consumer and business decisions. (…) lower real interest rates then lead to rises in business fixed investment, residential housing investment, inventory investment and consumer durable expenditure, (…)." | Mishkin et al. [101, p. 537] |

**Table 3:** continued

| Arrow | Causality assumption | Source |
|---|---|---|
| 47 | Investors face a choice between domestic and foreign assets and choose the investment with the highest expected return. | Blanchard et al. [102, p. 119] |
| 48 | The nominal exchange rate is fundamental to the determination of the real exchange rate. | Blanchard et al. [102, p. 112] |
| 49 | The degree of CBI plays a meaningful role only if the central bank places a different emphasis on alternative policy objectives than the government. The literature points to two main differences. One relates to possible differences between the rate of time preference of political authorities and that of central banks. For various reasons, central banks are often more conservative and take a longer view of the policy process than do politicians. The other difference concerns the subjective weights in the objective function of the central bank and that of the government. It is often assumed that central bankers are more concerned about inflation than about policy goals such as the achievement of high employment levels and adequate government revenues. | Eijffinger and de Haan [112, p. 7] |
| 50 | Central banks publicly communicate their inflation target or range. One benefit of IT adoption is "a well-known and credible inflation target helps to anchor the private sector's long-run inflation expectations." | Svensson [113, p. 1248] |
| 51 | Central banks publicly communicate when they peg their currency, which affects their credibility. | Burda and Wyplosz [99, pp. 492 f.] |
| 52 | Central bank operations affect the money supply. | Mishkin et al. [101, pp. 301 ff.] |
| 53 | The government's flow budget constraint means that current government debt is dependent on past debt (and other items). | Burnside [109, p. 36] |
| 54 | Given the breakdown of the relationship between monetary aggregates and goal variables such as inflation, many countries have recently adopted inflation targeting as their monetary policy regime. | Mishkin [114, pp. 590 f.] |
| 55 | Targeting the exchange rate is a monetary policy regime with a long history of adoption by central banks. | Mishkin [114, p. 581] |
| 56 | Monetary authorities react to changes in the demand for money. | Burda and Wyplosz [99, pp. 216–7] |
| 57 | Demand for the monetary base M0 (money produced by the central bank) depends negatively on the nominal interest rate. | Burda and Wyplosz [99, p. 217] |
| 58 | Demand for the monetary base M0 depends positively on nominal GDP. | Burda and Wyplosz [99, p. 217] |
| 59 | Masciandaro and Romelli (2019) found that more stable governments are associated with higher levels of CBI. Cukierman and Webb (1995) showed that CBI is lower in less stable political systems | Cukierman and Webb [57, p. 217] and Masciandaro and Romelli [54] |
| 60 | "When stock prices rise, the value of financial wealth increases, thereby increasing the lifetime resources of consumers, and consumption should rise." | Mishkin et al. [101, p. 542] |
| 61 | Savings lead to higher wealth. | Cooper and Dynan [115] |
| 62 | "[…], when monetary policy is expansionary, the public finds that it has more money than it wants and so gets rid of it through spending. One place the public spends is in the stock market, increasing the demand for stocks and consequently raising their prices." | Mishkin et al. [101, p. 542] |

**Table 3:** continued

| Arrow | Causality assumption | Source |
|---|---|---|
| 63 | Tobin defines *q* as the market value of firms divided by the replacement cost of capital. | Mishkin et al. [101, p. 540] |
| 64 | Asset returns have a significant effect on household savings. | Disney et al. [116] |
| 65 | Central banks' main objective is stable and low inflation. When inflation exceeds, or is expected to exceed, a certain level, a reaction by the central bank follows. | Taylor [117] |
| 66 | "[…] the bank lending channel of monetary transmission operates as follows: expansionary monetary policy, which increases bank reserves and bank deposits, increases the quantity of bank loans available." | Mishkin et al. [101, pp. 542 f.] |
| 67 | "Because many borrowers are dependent on bank loans to finance their activities, this increase in loans will cause investment (and possibly consumer) spending to rise […]." | Mishkin et al. [101, pp. 542 f.] |
| 68 | Central bank transparency is multidimensional, covering political transparency (openness about policy objectives), economic transparency (openness about data, models, and forecasts), procedural transparency (openness about the way decisions are made, achieved mainly through the release of minutes and votes), policy transparency (openness about the policy implications, achieved through prompt announcement and explanation of decisions), and operational transparency (openness about the implementation of those decisions). Transparency is a means of enhancing the credibility of central bank commitments. | Dincer and Eichengreen [45] |
| 69 | The most prominent argument for CBI is based on the time-inconsistency problem. It arises when the best plan made in the present for some future period is no longer optimal when that period actually starts. Implicitly, CBI reduces the time preference of monetary policy makers. | Eijffinger and de Haan [112, p. 5] |
| 70 | When a country becomes more open in economic terms, the nontraded sector becomes less important than the traded goods sector. | Lane [118] |
| 71 | The more important the traded good sector is, the less that monetary authorities stand to gain from surprise inflation because a monetary expansion in an open economy will be accompanied by a real depreciation of the currency, raising costs for households and businesses. The larger the share of imported goods is, the greater the increase in inflation. | Lane [118] and Romer [119] |
| 72 | *Past Inflation* can be considered as a summary statistics of past consumer price movements. | By definition. |
| 73 | "The hybrid Phillips curve is an example of how models used in the policy arena seek to overcome unsatisfactory features of both the adaptive expectations Phillips curve (it is empirically successful, but is subject to the Lucas critique; lacks micro-foundations and rational expectations; and lacks a channel for credibility to affect inflation) and the NKPC (which is forward looking and therefore not subject to the Lucas critique; has micro-foundations and rational expectations with a role for credibility, but counterfactual empirical predictions). The hybrid Phillips includes forward-looking inflation | Carlin and Soskice [120, p. 610] |

**Table 3:** continued

| Arrow | Causality assumption | Source |
|---|---|---|
| | expectations but acknowledges that inflation appears to be persistent or inertial, i.e., that it depends on lagged values of itself […]. The hybrid Phillips curve can be rationalized by the assumption that some proportion of firms use a backward-looking rule of thumb to set their inflation expectations while the remainder use forward-looking expectations." | |
| 74 | One way for a central bank to establish credibility is by increasing its independence. | Blinder [121] |
| 75 | Employees want to protect themselves from a loss in purchasing power, so they embed their inflation expectations into their nominal wages. | Burda and Wyplosz [99, p. 293] |
| 76 | "Expansionary monetary policy, which causes a rise in stock prices along the lines described earlier, raises the net worth of firms […]" | Mishkin et al. [101, p. 544] |
| 77 | "The lower the net worth of business firms, the more severe the adverse selection and moral hazard problems in lending to these firms. Lower net worth means that lenders in effect have less collateral for their loans, so their potential losses from adverse selection are higher." | Mishkin et al. [101, p. 544] |
| 78 | "The lower net worth of businesses also increases the moral hazard problem because it means that owners have a lower equity stake in their firms, giving them more incentive to engage in risky investment projects. Because taking on riskier investment project makes it more likely that lenders will not be paid back, a decrease in businesses' net worth leads to a decrease in lending and hence in investment spending." | Mishkin et al. [101, p. 544] |
| 79 | In a more integrated world, competition between currencies is even more present since countries want to attract foreign investments, and this race is exacerbated in a financially integrated world. | Wei and Tytell [122] |
| 80 | The primary mechanism through which greater openness to foreign capital might lead to lower inflation is presumably some sort of disciplining effect on monetary policy. | Wei and Tytell [122] |
| 81 | The quality of political institutions might directly influence the relationship between CBI and inflation. The effectiveness of CBI in strengthening credibility and enhancing inflation performance is increased by the presence of multiple political veto players or if checks and balances are sufficiently strong. | Keefer and Stasavage [123] and Hayo and Voigt [124] |
| 82 | Political instability can have a number of possible effects. The most commonly discussed of these is that more instability makes it difficult for policy makers to commit to low inflation. | Campillo and Miron [125, p. 10] |
| 83 | Income per capita captures several possible effects. A higher level of income per capita is likely to be accompanied by a more sophisticated tax system and a more developed financial system, both of which imply a lower optimal inflation tax and thus a negative relation with inflation. On the other hand, high-income countries might be better at innovating technologies for reducing the costs of inflation, so their inflation aversion might be lower. | Campillo and Miron [125, p. 11] |

**Table 3:** continued

| Arrow | Causality assumption | Source |
|---|---|---|
| 84 | The life-cycle theory suggests that individuals plan their consumption and savings behavior over their life-cycle and smooth out their consumption over their lifetimes. Aggregate demand and supply shift because certain age groups and their particular economic behavior gain in relative importance to the rest of the population. Hence, changes in the demographic structure can exert potentially large effects on total savings. | Bobeica et al. [126, p. 5] |
| 85 | For given prices, nominal and real interest rates are directly linked through the Fisher equation. | Burda and Wyplosz [99, p. 524] |
| 86 | The government introduces or changes the law of the central bank that affects its obligations to provide more transparency. | By definition |
| 87 | "Another balance sheet channel operates by affecting cash flow, the difference between cash receipts and cash expenditures. The rise in cash flow increases the liquidity of the firm (or household) and thus makes it easier for lenders to know whether the firm (or household) will be able to pay its bills. The result is that adverse selection and moral hazard problems become less severe, […]" | Mishkin et al. [101, p. 544 f.] |
| 88 | Money demand depends on nominal output, so the price level becomes relevant. | Burda and Wyplosz [99, p. 217] |
| 89 & 90 | The government also collects its revenue through tax payments. | Walsh [106, p. 136] |
| 91 | If the government runs a budget deficit by holding spending constant and reducing tax revenue, households' current disposable income, and perhaps their lifetime wealth, increases. | Elmendorf and Mankiw [127] |
| 92 | If the exchange-rate target is credible, it anchors inflation expectations to the inflation rate in the anchor country to whose currency it is pegged. | Mishkin [114, p. 581] |
| 93 | Pegging the exchange rate to a foreign anchor forces the country to adopt the foreign interest rate policy, which affects broad domestic money supply. | Mishkin [114] |
| 94 | In an inflation (forecast) targeting framework, the central bank changes its short-term interest rate if the inflation forecast exceeds or falls short of the inflation target, until the inflation forecast equals the target. In a related version of the inflation targeting strategy, the central bank may deem it appropriate to adjust its monetary policy if the inflation forecast indicates a deviation from target (or its range). In either case, the money supply will be affected. | Svensson[128] and and Jordan et al. [129] |
| 95 | A government that issues nominal debt has an incentive to promise low inflation *ex ante* to lower nominal interest payments and then reduce the ex post value of the debt through unexpected inflation. This incentive is stronger the larger the public debt is. | Kwon et al. [130] |
| 96 | Credit growth is a more important determinant of consumption than income growth. | [131] |
| 97 | Capital and savings are usually valued by discounting. The amount of the discount depends primarily on the real interest rate. | Burda and Wyplosz [99, p. 161] |
| 98 | The government legislates on the degrees of independence to be granted to the central bank. | By definition |

**Table 3:** continued

| Arrow | Causality assumption | Source |
|---|---|---|
| 99 | Governments in federally organized countries with good checks and balances grant their monetary institutions greater autonomy (De Haan and van't Hag (1995), Moser (1999), Keefer and Stasavage (1999), or Farvaque (2002)). Reforms are also more likely following elections that lead to a political consolidation or to changes in the political orientation of the government (Romelli, 2018). Masciandaro and Romelli (2019) found evidence that countries in common-law jurisdictions have a lower degree of independence. | Farvaque [47], Romelli [53], Masciandaro and Romelli [54], de Haan and van 't Hag [55], Moser [56], Keefer and Stasavage [123] |
| 100 | Binding agreements with international lenders like the IMF or the World Bank often require countries to commit to a particular set of policies (Blejer et al., 2002; Gutierrez, 2003; Polillo and Guillén, 2005; Rodrik and Bank, 2006; Romelli, 2018; Kern, Reinsberg and Rau-Göhring, 2019; Reinsberg et al. (2020)). Another type of external pressure can come from regional clustering, which is often found to be cohesive of certain types of reform processes such as democratizations and economic liberalizations (Simmons and Elkins, 2004; Elhorst et al., 2013; Giuliano et al., 2013; Acemoglu et al., 2019). | Kern et al. [43], Polillo and Guillén [49], Romelli [53], Blejer et al. [59], Gutiérrez [60], Rodrik [61], Reinsberg et al. [62], Simmons and Elkins [63], Elhorst et al. [64], Giuliano et al. [65], Acemoglu et al. [66] |
| 101 | Masciandaro and Romelli (2019) found that countries experiencing high inflation in the past have a higher inflation aversion which constrains the government to assign a higher degree of CBI in the following periods. Similarly, Crowe and Meade (2008) showed that over the period 1990–2003, greater changes in CBI have occurred in countries with higher inflation. According to Wachtel and Blejer (2020), the arguments in favor of an independent central bank began to crystallize in the 1980s after a decade or more of traumatic inflationary experience that put a spotlight on central bank policymaking and its failures. | Crowe and Meade [13], Masciandaro and Romelli [54], Wachtel and Blejer [58] |

# B Additional material related to the data analysis

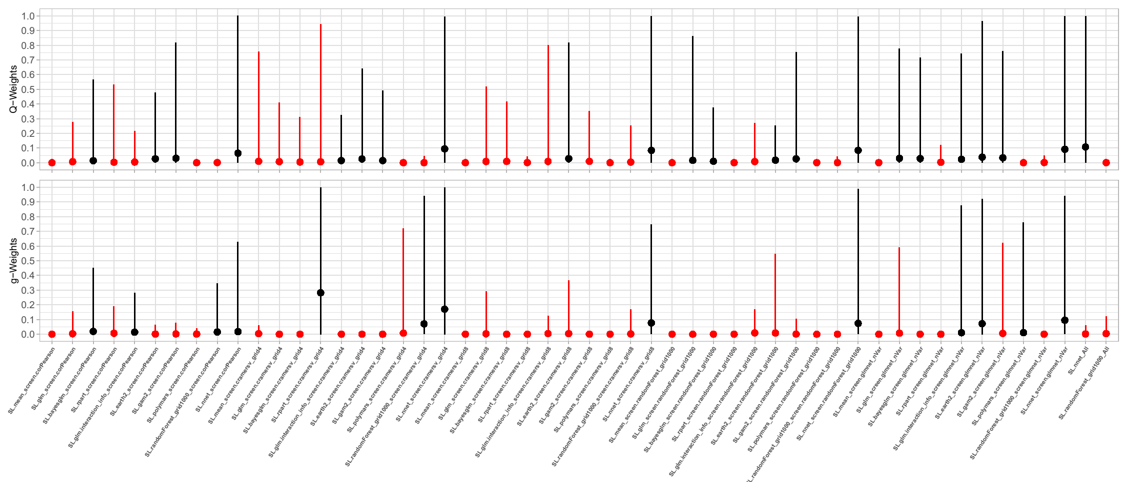


**Figure 4:** Distribution of learner weights. The visualized distributions are based on the merged learner weights that resulted from the estimation of $\Psi_{1,3}$ and $\Psi_{2,3}$ ($\bar{d}_{t^*}^1$, $\bar{d}_{t^*}^2$, and twice $\bar{d}_{t^*}^3$), summarized across the imputed data sets. The plotted point represents the mean of each distribution. If it is below 0.01, both the distribution and the mean are displayed in red.

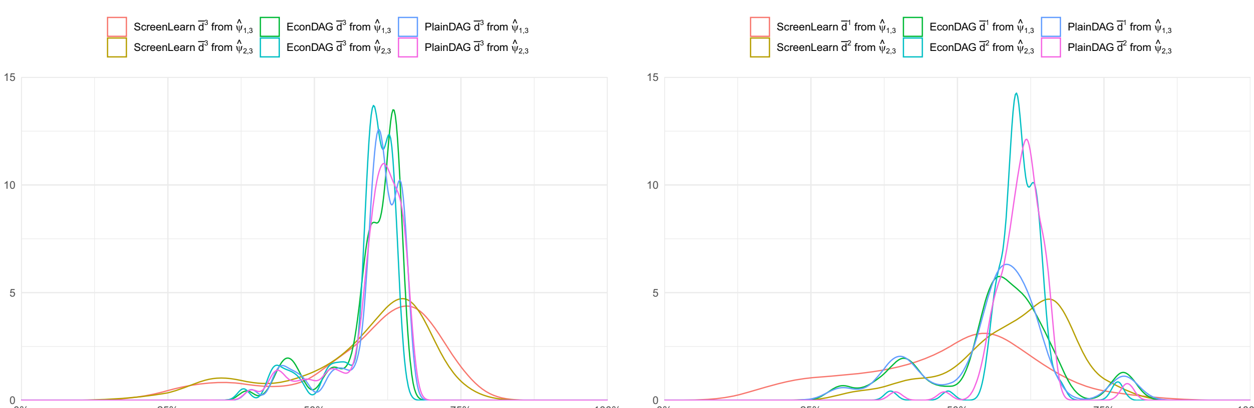


**Figure 5:** Kernel density plots of cumulative treatment probabilities for $T = 2010$. In the left panel, estimated probabilities of $\bar{d}_{t^*}^3$ are shown while the right panel shows estimated probabilities for $\bar{d}_{t^*}^1$ and $\bar{d}_{t^*}^2$.

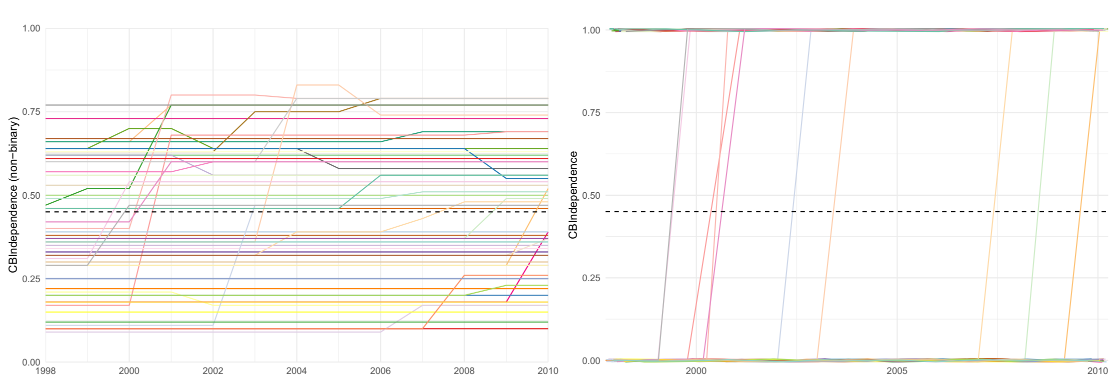


**Figure 6:** Trajectories of the intervention variable (CBI) for all included countries ($n$ = 60).

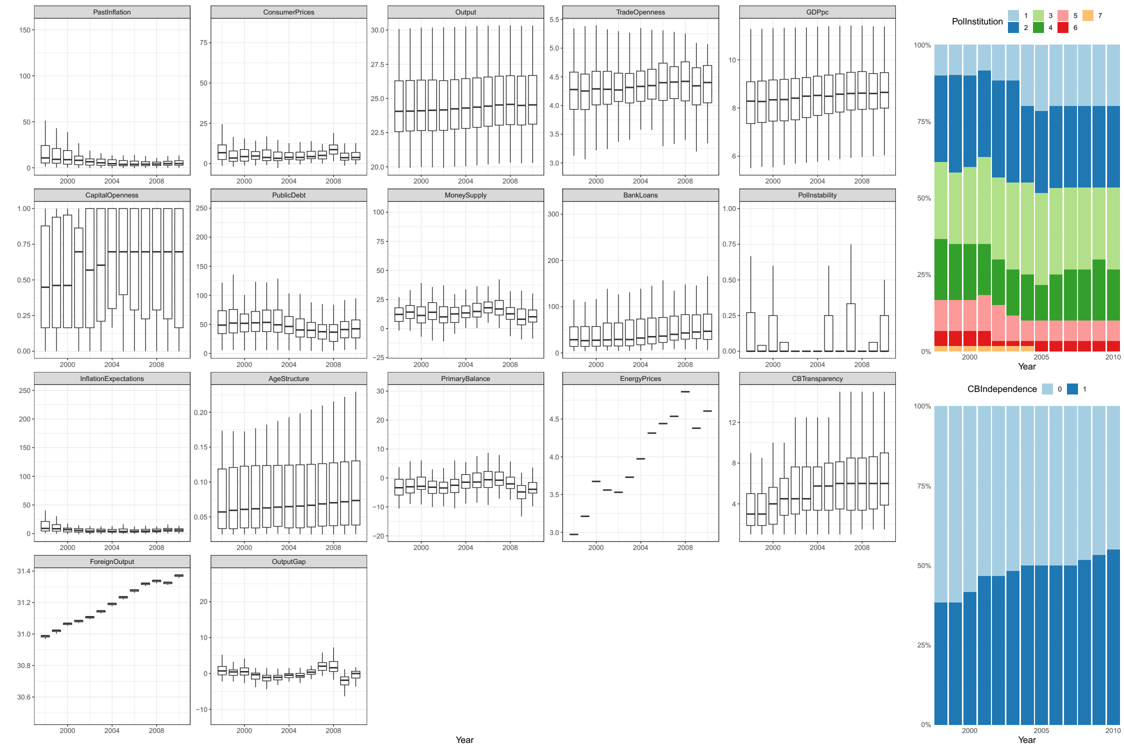


**Figure 7:** Summary statistics for all variables included in the data analysis. The variables “Output,” “TradeOpenness,” “GDPpc,” “EnergyPrices,” and “ForeignOutput” were transformed by the natural logarithm for better readability. Appendix A gives more details on the variables and what they measure.

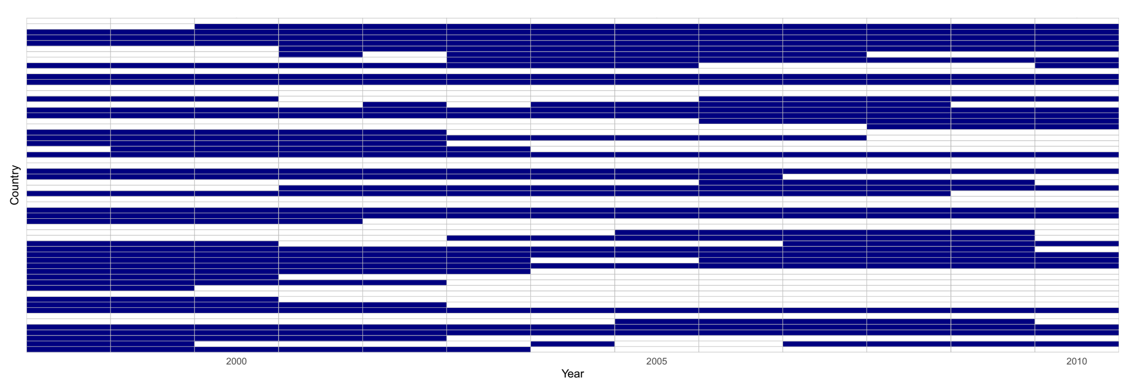


**Figure 8:** Blue tiles indicate when a country has had a negative or above 5% median inflation rate in the last 7 years while having legislated an independent central bank simultaneously.

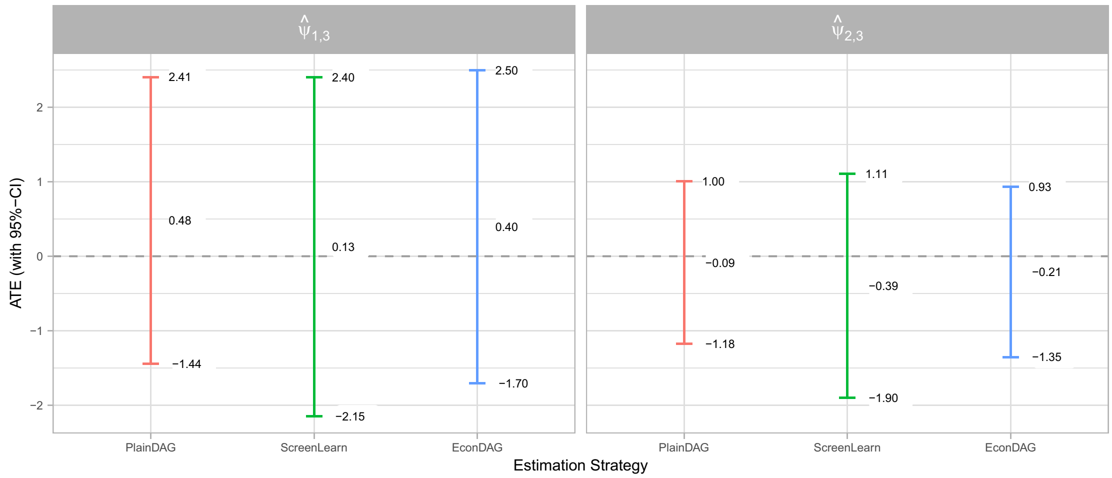


**Figure 9:** ATE among the high-income countries ($n$ = 26).

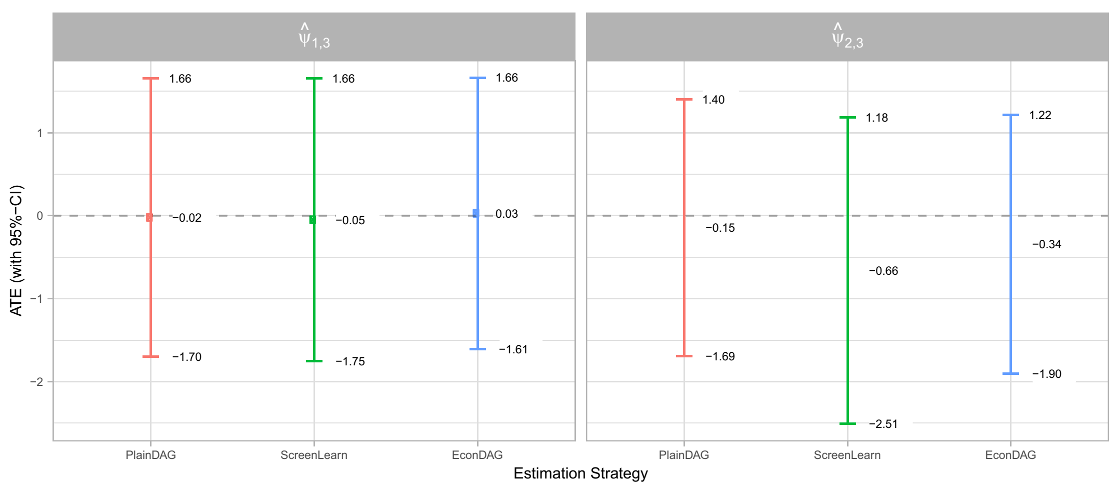


**Figure 10:** ATE among the low-income countries ($n$ = 34).

# C Details on the simulation study

## C.1 IPTWC.2 Data-generating processes (DGP)

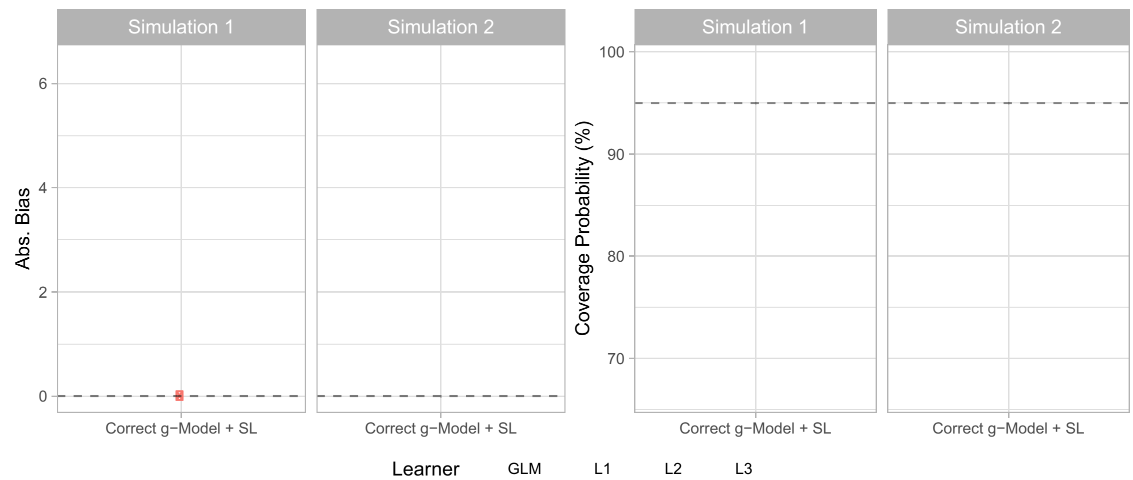


**Figure 11:** Absolute bias and coverage probabilities for estimation with IPTW. Bias: 0.009 (GLM), 6.377 (L1), 6.325 (L2), 6.431 (L3) and coverage probability: 99.1% (GLM), 67.3% (L1), 66.6% (L2), 66.3% (L3).

**Table 4:** DGP for simulation 1

| $t = 1$ | $t = 2, 3$ |
|---|---|
| $L_t \sim N(0, 0.25)$ | $L_t \sim N(L_{t-1} + A_{t-1}, 0.25)$ |
| $A_t \sim B(\text{expit}(L_t))$ | $A_t \sim B(\text{expit}(L_t + 2 \times A_{t-1} - L_{t-1}))$ |
| $Y_t \sim N(50 \times A_t + L_t, 0.25)$ | $Y_t \sim N(50 \times A_t + L_t + L_{t-1} + Y_{t-1}, 0.06)$ |

**Table 5:** DGP for simulation 2

| $t = 1$ | $t = 2, \dots, 6$ |
|---|---|
| $L_t^1 \sim N(0, 0.25)$ | $L_t^1 \sim N(L_{t-1}^7, 0.25)$ |
| $A_t \sim B(\text{expit}(L_t^1))$ | $A_t \sim B(\text{expit}(0.25 \times L_t^1 + 0.25 \times L_{t-1}^6))$ |
| $Y_t \sim N(A_t + L_t^1, 9)$ | $Y_t \sim N(A_t + L_t^1 + L_{t-1}^9 + 0.05 \times L_{t-1}^{10}, 0.25)$ |
| $L_t^5 \sim N(Y_t, 2.25)$ | |
| $t = 1, \dots, 5$ | $t = 1, \dots, 5$ |
| $L_t^2 \sim N(A_t + L_t^1, 0.25)$ | $L_t^5 \sim N(Y_t + L_{t-1}^{10}, 2.25)$ |
| $L_t^3 \sim N(Y_t + L_t^2, 1)$ | |
| $L_t^4 \sim N(A_t, 0.25)$ | |
| $L_t^6 \sim N(L_t^4, 0.25)$ | |
| $L_t^7 \sim N(L_t^2, 0.25)$ | |
| $L_t^8 \sim N(L_t^5, 0.25)$ | |
| $L_t^9 \sim N(L_t^3, 1)$ | |
| $L_t^{10} \sim N(L_t^8 + L_t^9, 0.25)$ | |