A model for policy interest rates

Armin Seibert, Andrei Sirchenko, Gernot Müller

A model for policy interest rates☆

Armin Seibert [a], Andrei Sirchenko [b,*], Gernot Müller [a]

[a] *Augsburg University, Universitaetsstrasse 14, 86159 Augsburg, Germany*
[b] *University of Amsterdam, P.O. Box 15867, 1001 NJ Amsterdam, the Netherlands*

A B S T R A C T

This paper introduces a model that addresses the key worldwide features of modern monetary policy making: the discreteness of policy interest rates both in magnitude and in timing, the preponderance of status quo decisions, monetary policy inertia and policy regime switching. We capture them by developing a new dynamic ordered-choice model with switching among three latent policy regimes (easing, neutral and tightening). The simulations and an application to the federal funds rate target demonstrate that ignoring these features leads to biased estimates, worse in- and out-of-sample predictions, and a qualitatively different inference. Using all Federal Open Market Committee's (FOMC) decisions made both at scheduled and unscheduled meetings as sample observations, we model the Federal Reserve's response to real-time data available right before each meeting. The new model, fitted for the Greenspan's tenure, detects oscillating switches among latent regimes, identifies three types of status quo decisions, correctly predicts out of sample 90% of the next 111 FOMC decisions on the target rate, and clearly outperforms the linear models (including the Taylor rule), the conventional ordered probit and other discrete-choice models from the literature.

1. Introduction

We develop a dynamic model of discrete ordered choice with lagged latent dependent variables among regressors and with time-varying transition probabilities of switching among three latent regimes interpreted in the interest-rate-setting context as easing, neutral and tight monetary policy stances. The model assumes three implicit decisions: the policy regime decision and two decisions (conditional on the easing or tight regime) on the amount of rate change. The new methodology synthesizes and extends the existing models of ordered choice and avoids important common distortions of the data-generating process (DGP) in the empirical identification of monetary policy rules.

The policy interest rates are currently a critical instrument and a principal measure of monetary policy in many countries. They are an anchor for other short-term market interest rates, and are closely watched and anticipated by many economic

agents. If we can quantitatively formalize how monetary authorities set the policy rates, and "if practitioners in financial markets gain a better understanding of how policy is likely to respond to incoming information, asset prices and bond yields will tend to respond to economic data in ways that further the central bank's policy objectives" (Bernanke, 2007). To forecast the state of the economy, or to evaluate the effects of economic shocks, monetary and fiscal policy actions, we need to understand central bank's systematic response to economic data. To improve the monetary policy, we also need a clear empirical description of what is going to be improved. It is difficult to describe monetary policy without using an econometric model. Since the policy rates are set administratively by the monetary authorities, and are neither the outcomes of market interaction of supply and demand nor subject to technical fluctuations or extraneous sources of noise, it makes them of special interest for econometric modeling (Hamilton and Jorda, 2002).

We implement an econometric framework that addresses the main worldwide features of modern monetary policy making: the discreteness of policy interest rates, both in magnitude and in timing; the preponderance and heterogeneity of status quo (no change to the rate) decisions; interest rate smoothing and policy inertia; policy regime switching and different nature of positive and negative interest rate movements. None of the existing models is able to adequately capture all these stylized facts. We extend and synthesize the models of Brooks et al. (2012), Dale and Sirchenko (2021), Hamilton and Jorda (2002), Harris and Zhao (2007), Müller and Czado (2005), and Sirchenko (2019, 2020). More specifically, we address the following issues.

The discreteness of policy rates. Monetary policy modeling has been historically implemented in a continuous framework, either by estimating a 'simple' linear policy rule such as the Taylor rule (Taylor, 1993) or by estimating a monetary policy reaction function as a linear equation of a vector autoregressive model.[1] Regression methods for a continuous dependent variable have always been and still are absolutely appropriate to model the market interest rates, both the long- and short-term ones, but not appropriate any longer for policy interest rates set by many central banks. Nowadays, many central banks (and all the major ones) change policy rates by discrete amounts, typically by multiples of 25 basis points (bp), at special meetings of monetary policy committees, which are held 6–12 times a year. The notable examples are the Federal Open Market Committee of the U.S. Federal Reserve System (Fed), the Governing Council of the European Central Bank, the Monetary Policy Committee of the Bank of England, and the Policy Board of the Bank of Japan, among others.

Despite the extensive literature on monetary policy modeling, most studies use monthly or longer-period data averages and employ regression techniques for a continuous dependent variable, and thus do not address the discreteness of interest rates. Ignoring the limited (discrete) nature of the dependent variable may lead to serious biases. While using regressions for continuous outcomes is appropriate with aggregated data, it raises the problems of distortions caused by the use of an incorrect information set and of simultaneity caused by disregarding possible interactions between the policy rates and the other variables that can happen during a period of aggregation. In recent years, discrete-choice approaches such as the ordered probit models (Gerlach, 2007; Hayo and Neuenkirch, 2010; Vanderhart, 2000) have been employed. There have been also attempts to use the discrete-choice models with FOMC-meeting data frequency (Hu and Phillips, 2004; Kauppi, 2012; Kim et al., 2009; Piazzesi, 2005); they, however, restricted their attention to the decisions made at the scheduled FOMC meetings only, thereby not reflecting the entire policy-making process. In fact, from 1983 to 1993 only about 15% of the changes to the U.S. federal funds rate target (*target* henceforth) have been made at the scheduled FOMC meetings.

We address the discrete nature of policy rates by an ordered choice approach. We carefully mimic the policy-making process by using all interest rate decisions (made both at the scheduled and unscheduled FOMC meetings) as sample observations. We match the FOMC decisions with the latest real-time vintages of macroeconomic and non-aggregated daily financial data truly available immediately before each meeting and not revised later on. Information about the exact timing of FOMC meetings together with the use of daily financial data (such as spreads between the long- and short-term market interest rates) allows us to substantially improve the identification of the Fed policy rule, which would not be possible to do using monthly aggregates of financial data. In addition, the market interest rates encapsulate the huge volume of data available to market participants, and can help avoid the omitted variables and time-varying parameter problems. The empirical results show that discreteness does matter in the estimation of monetary policy rules. The out-of-sample performance of the proposed model is remarkably superior to that of the linear model estimated using the same set of regressors: for example, the mean absolute forecast error is more than twice as small as in the linear model, and the percentage of correct predictions is higher by 19 percentage points.

The preponderance of status quo decisions. The central banks often prefer to wait and see — many of them leave the policy rates unchanged at more than a half of policy meetings. The central banks make status quo (no change) decisions in different macroeconomic circumstances: between rate hikes, between cuts and also between changes in different directions. See Fig. 1 for the FOMC decisions on the target changes during the 7/1987–1/2006 period.

Many of the status quo decisions observed between cuts are made in rather different economic circumstances than those observed between hikes. Many of the status quo decisions observed between changes in different directions (that is before rate reversals) are generated by economic conditions that differ a lot from those observed in the periods of unidirectional changes to the target. We address the heterogeneity of status quo decisions by a three-part modelling approach that allows status quo outcomes to be generated by three different latent regimes interpreted as easing, neutral and tight policy stances. We extend the zero-inflated model of Harris and Zhao (2007) and the middle-inflated model of Brooks et al. (2012) by

---

[1] See Rudebusch (1998a,b) for a critique of monetary policy identification in the context of vector autoregressive models.
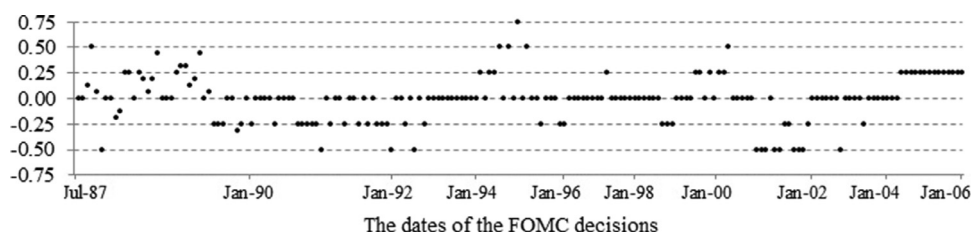
Fig. 1. Historical changes to the federal funds rate target during the Greenspan tenure.

making them suitable for ordinal outcomes that take on negative, zero and positive values, and by allowing them to be past-dependent and autocorrelated.

We detect the oscillating switches among three latent regimes during the relatively stable period such as the Greenspan's tenure, identify three types of the status quo decisions, and find out that almost a half of them are generated in either easing or tight policy regime. In an out-of-sample forecasting exercise, during the zero-lower bound (ZLB) period, the proposed model correctly predicts 53 out of 55 status quo decisions, successfully treating the ZLB by an extended switch to an easing policy regime. By contrast, the Taylor rule and linear model fail to manage the ZLB and correctly forecast only 34 no-change decisions, wrongly forecasting 21 further cuts.

Regime switching. The central banks may have asymmetric responses to incoming data, so the rate increases and decreases may be generated by different decision-making processes as well as the numerous status quo decisions may also be driven by distinct mechanisms. In other words, the policy actions may be generated in different regimes.

A change to the monetary policy regime is often modelled as a permanent, non-stochastic switch: empirical studies divide samples into regime-specific subperiods or employ a gradual regime switching approach with smoothed transitions between regimes. However, treating regime changes as breaks rather than as a slowly fluctuating stochastic process is largely a matter of convenience and even logically inconsistent, as Cooley et al. (1984) argue. The literature on modeling short-lived oscillating switches in monetary policy demonstrate that dividing samples into regime-specific subperiods can distort qualitative and quantitative inferences, and that monetary policy changes are better modeled as stochastic and repeated fluctuations among two or more regimes — see, among others, Assenmacher-Wesche (2006) in the univariate context, Bikbov and Chernov (2013) and Sims and Zha (2006) in the structural vector autoregression context, and Andolfatto and Gomme (2003), Bianchi (2013), Davig and Leeper (2007) and Baele et al. (2015) in the dynamic stochastic general equilibrium (DSGE) framework.

Despite the increasing popularity of a regime-switching approach in macroeconomics, the existing regime-switching applications in monetary economics focus on models for continuous outcomes. Building on the seminal work of Hamilton (1989), recurring policy regimes are typically modelled as the unobserved states generated by a stochastic exogenous Markov-chain process, which is independent of the endogenous economic variables and has constant probabilities of transition from one regime to another. This assumption is realistic in a wide range of contexts but highly unlikely in monetary policy applications. It is natural to assume that central banks react systematically to changes in the economic conditions; thus, regime switches are endogenous to the state of the economy and have the time-varying transition probabilities.

We model this possibility by a regime-switching approach with time-varying transition probabilities of switching among three latent regimes. It allows the probabilities of positive, negative and no changes to be treated differently and to be asymmetrically affected by the economic data. The structure of the new model shares some similar features with the Markov switching models with constant probabilities of a transition from one regime to another introduced by Hamilton (1989) in the framework of linear regressions for a continuous dependent variable, later extended by Diebold et al. (1994), Filardo (1994) and Bazzi et al. (2017) to the case when the transition probabilities may change over time. We generalize these endogenous switching models for continuous outcomes to discrete outcomes.

Interest rate smoothing. The policy interest rates are persistent. This stylized fact is usually referred to as interest rate smoothing. If central banks decide to change the rate they will most likely move it again in the same direction several times at the next meetings, avoiding frequent rate reversals. This feature is referred to as monetary policy inertia. We address these two stylized facts by the lagged latent dependent variable introduced to the covariates adopting the autoregressive ordered probit (AOP) model of Müller and Czado (2005) in each latent equation. It allows for a partial adjustment of interest rates and together with the lagged covariates can capture the autocorrelation of latent monetary shocks.

The dynamic single-equation ordered probit models have been developed and applied to policy interest rates, among others, by Davutyan and Parke (1995), Dueker (1999), Eichengreen et al. (1985), Hu and Phillips (2004), Kim et al. (2009), and Monokroussos (2011). Hamilton and Jorda (2002) developed the dynamic two-equation ordered probit model, in which the first-stage binary decision (change or no change) is determined by the autoregressive conditional hazard model, and the magnitude of rate changes is determined by the ordered probit model conditional on a change at the first stage. Grammig and Kehrle (2008) modified this model by implementing the autoregressive conditional multinomial model of Russel and Engle (2005) at the second stage. Each stage in these models is estimated separately, and status quo observa-

tions are not included in the estimation of the second stage. We further extend these two-part models by implementing at the first stage a trichotomous decision on the latent policy regime (easing, neutral or tight) that seems to be more realistic than a binary decision (change or no change): the policymakers, who are determined to change the rate, have already chosen the direction of change. Furthermore, we estimate both stages simultaneously and do not exclude no-change outcomes from the second stage. It allows us to discriminate among different types of the status quo decisions.

In the next section we describe the proposed Switching Autoregressive Ordered Probit (SwAOP) model. We discuss the estimation and inference in Section 3. The estimation of such a dynamic three-equation ordered probit model is a daunting computational challenge − it requires the evaluation of multiple integrals with no closed form solution. We opt for a Bayesian approach using Markov chain Monte Carlo (MCMC) methods (a Gibbs sampler), which makes estimation computationally feasible and requires no numerical optimization and no high-dimensional integration. In Section 4 we study the finite sample performance of the SwAOP estimator under both its own DGP and the AOP one. In Section 5 we discuss the empirical application to the FOMC decisions on the target. Section 6 concludes. Supporting materials are provided in the Appendix A and Online Appendix B.

2. Model

The SwAOP model is designed for changes in ordinal variables, which are subject to a regime switching or a sample separation and have abundant and heterogeneous observations in a middle neutral category (typically, a zero or no-change). We illustrate the model in the context of monetary authority decisions on the policy interest rate. The observed dependent variable is a change to the policy rate $\Delta y_t = y_t - y_{t-1}$, where y_t is the level of the rate set at the meeting $t = 1, 2, ., T$; Δy_t takes on a finite number of discrete ordered values j (typically, in steps of 25 basis points (bp)) coded as $J^-, \ldots, -1, 0, 1, \ldots, J^+$. We mark the latent (unobserved) variables by an asterisk (*). Fig. 2 shows the decision tree of the model.

The model assumes three latent regimes, which are observed only partially, namely, only if the observed outcome is a change; if no change is observed, the regime is unknown. The regimes are determined by the continuous latent variable r_t^{0*} representing the degree of central bank policy stance according to a latent *regime decision*

$$r_t^{0*} = \phi^0 r_{t-1}^{0*} + \beta^0 \mathbf{x}_t^0 + \varepsilon_t^0, \tag{1}$$

where $\boldsymbol{\beta}^0$ is a vector of k^0 unknown slope parameters, $\mathbf{x}_t^0$ is the tth column of the observed $k^0 \times T$ data matrix $\mathbf{X}^0$, ϕ^0 is an unknown autoregressive (AR) parameter, and ε_t^0 is an error term that is independently and identically distributed (IID) according to the standard normal cumulative distribution function (CDF) Φ.

The observed discrete change Δy_t is conditional on a latent discrete variable s_t^* (coded as -1, 0, or 1 if the central bank policy regime is easing, neutral or tight, respectively). Both Δy_t and s_t^* are determined in an ordered probit fashion as

$$
\begin{array}{lll}
\Delta y_t | (s_t^* = -1) = j & \text{if} & c_{j-1}^- < r_t^{-*} \leq c_j^- \text{ for } j \leq 0, \\
\Delta y_t | (s_t^* = 0) = 0, & & \\
\Delta y_t | (s_t^* = 1) = j & \text{if} & c_j^+ < r_t^{+*} \leq c_{j+1}^+ \text{ for } j \geq 0,
\end{array}
\qquad
s_t^* = \begin{cases} -1 & \text{if} \quad r_t^{0*} \leq c_1^0, \\ 0 & \text{if} \quad c_1^0 < r_t^{0*} \leq c_2^0, \\ 1 & \text{if} \quad c_2^0 < r_t^{0*}, \end{cases}
$$

where $-\infty < c_1^0 \leq c_2^0 < \infty$, $-\infty \equiv c_{J-1}^- \leq c_{J-}^- \leq \ldots \leq c_0^- \equiv \infty$ and $-\infty \equiv c_0^+ \leq c_1^+ \leq \ldots \leq c_{J+1}^+ \equiv \infty$ are the unobserved cut-point parameters; and r_t^{-*} and r_t^{+*} are the continuous latent variables, representing the desired amount of the rate change
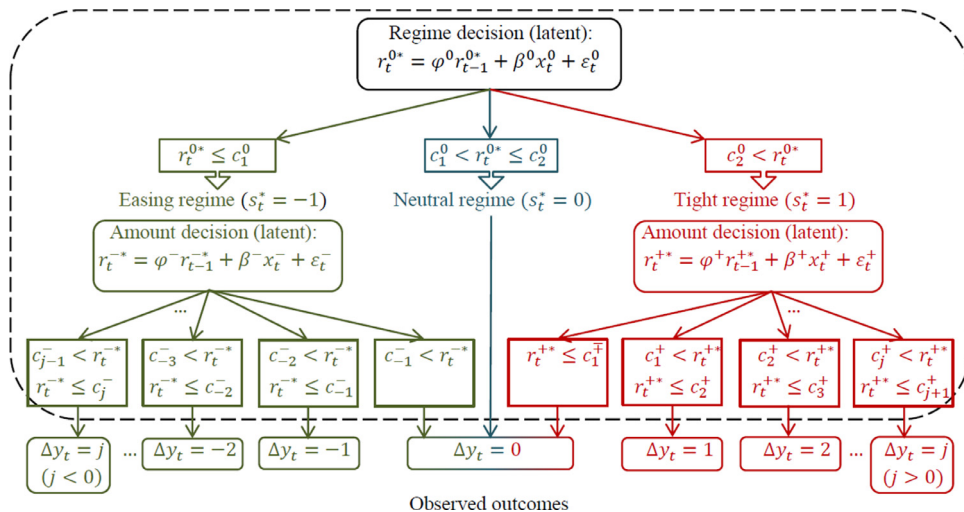

Fig. 2. Decision tree of the SwAOP model.

in the easing and tight regimes, respectively. They are driven by the latent *amount decisions*

$$r_t^{-*} = \phi^- r_{t-1}^{-*} + \beta^- \mathbf{x}_t^- + \varepsilon_t^- \qquad \text{and} \qquad r_t^{+*} = \phi^+ r_{t-1}^{+*} + \beta^+ \mathbf{x}_t^+ + \varepsilon_t^+, \tag{2}$$

where β^- and β^+ are the vectors of k^- and k^+ unknown slope parameters, $\mathbf{x}_t^-$ and $\mathbf{x}_t^+$ are the tth columns of the observed $k^- \times T$ and $k^+ \times T$ data matrices $\mathbf{X}^-$ and $\mathbf{X}^+$, ϕ^- and ϕ^+ are the unknown AR parameters, and ε_t^- and ε_t^+ are the IID error terms with the standard normal CDF Φ. The errors ε_t^0, ε_t^- and ε_t^+ are mutually independent.

Henceforth, the superscript indexes '0', ' - ' and ' + ' refer to the regime decision (1) and two amount decisions (2) conditional on the easing and tight regimes, respectively.

The SwAOP model can be summarized as

$$\Delta y_t = \begin{cases} j \ (j < 0) & \text{if} \quad r_t^{0*} \le c_1^0 \text{ and } c_{j-1}^- < r_t^{-*} \le c_j^-, \\ 0 & \text{if} \quad \begin{cases} c_1^0 < r_t^{0*} \le c_2^0, \\ \text{or } (r_t^{0*} \le c_1^0 \text{ and } c_{-1}^- < r_t^{-*}), \\ \text{or } (c_2^0 < r_t^{0*} \text{ and } r_t^{+*} \le c_1^+), \end{cases} \\ j \ (0 < j) & \text{if} \quad c_2^0 < r_t^{0*} \text{ and } c_j^+ < r_t^{+*} \le c_{j+1}^+. \end{cases}$$

To simplify the notation we let $e_{t,m}^i := c_m^i - \phi^i r_{t-1}^{i*} - \beta^i \mathbf{x}_t^i$ for $i \in \{0, -, +\}$. The probabilities to observe the outcome j are then given by

$$\Pr(\Delta y_t = j) = I_{j \le 0} \Phi\left(e_{t,1}^0\right)\left[\Phi\left(e_{t,j}^-\right) - \Phi\left(e_{t,j-1}^-\right)\right]$$
$$+ I_{j=0}\left[\Phi\left(e_{t,2}^0\right) - \Phi\left(e_{t,1}^0\right)\right] + I_{j \ge 0}\left[1 - \Phi\left(e_{t,2}^0\right)\right]\left[\Phi\left(e_{t,j+1}^+\right) - \Phi\left(e_{t,j}^+\right)\right], \tag{3}$$

where $I_{j \le 0}$, $I_{j=0}$ and $I_{j \ge 0}$ are the indicator functions such that: $I_{j \le 0} = 1$ if $j \le 0$ and $I_{j \le 0} = 0$ otherwise; $I_{j=0} = 1$ if $j = 0$ and $I_{j=0} = 0$ otherwise; and $I_{j \ge 0} = 1$ if $j \ge 0$ and $I_{j \ge 0} = 0$ otherwise.

As is common in latent-class models of discrete choice, the parameters are identified only jointly. The slope and autoregressive parameters are identified jointly with the variance of the error term; in particular, only their ratios ($\frac{\beta_{slopes}^0}{\text{Var}(\varepsilon_t^0)}$, $\frac{\phi^0}{\text{Var}(\varepsilon_t^0)}$, $\frac{\beta_{slopes}^-}{\text{Var}(\varepsilon_t^-)}$, $\frac{\phi^-}{\text{Var}(\varepsilon_t^-)}$, $\frac{\beta_{slopes}^+}{\text{Var}(\varepsilon_t^+)}$, $\frac{\phi^+}{\text{Var}(\varepsilon_t^+)}$) are identified. The cutpoint parameters are identified jointly with the intercept parameter and the variance of the error term; in particular, only the following quantities are identified: $c_j^0 - \frac{\beta_{intercept}^0}{\text{Var}(\varepsilon_t^0)}$, $c_j^- - \frac{\beta_{intercept}^-}{\text{Var}(\varepsilon_t^-)}$ and $c_j^+ - \frac{\beta_{intercept}^+}{\text{Var}(\varepsilon_t^+)}$. To identify the parameters we fix the variances of ε_t^0, ε_t^- and ε_t^+ to one, and the intercept components of β^0, β^- and β^+ to zero. These identifying assumptions are arbitrary and standard in the discrete-choice modeling. The probabilities in (3) are absolutely identifiable and invariant to the choice of parameter-identifying assumptions.

The marginal effect (ME) of the covariates on the probabilities is the partial derivative of the probabilities with respect to one of the covariates x_m, $ME_{m,j,t} := \partial_{x_{t,m}} \Pr(\Delta y_t = j)$ keeping all other parameters and covariates constant:

$$ME_{m,j,t} = I_{j=0}\left[f\left(e_{t,2}^0\right) - f\left(e_{t,1}^0\right)\right]\beta_m^{0,all}$$
$$+ I_{j \le 0}\left\{f\left(e_{t,1}^0\right)\left[\Phi\left(e_{t,j}^-\right) - \Phi\left(e_{t,j-1}^-\right)\right]\beta_m^{0,all} + \Phi\left(e_{t,1}^0\right)\left[f\left(e_{t,j}^-\right) - f\left(e_{t,j-1}^-\right)\right]\beta_m^{-,all}\right\}$$
$$+ I_{j \ge 0}\left\{f\left(e_{t,2}^0\right)\left[\Phi\left(e_{t,j+1}^+\right) - \Phi\left(e_{t,j}^+\right)\right]\beta_m^{0,all} + \Phi\left(-e_{t,2}^0\right)\left[f\left(e_{t,j+1}^+\right) - f\left(e_{t,j}^+\right)\right]\beta_m^{+,all}\right\},$$

where f is the normal probability density function (PDF), and $\beta^{i,all}$ is a vector of the slope coefficients in the latent equation i on all covariates in the model (in $\mathbf{X}^0$, $\mathbf{X}^-$ and $\mathbf{X}^+$): $\beta_m^{i,all} = 0$ if the covariate m does not appear in the equation i. We compute the MEs and probabilities at the empirical medians of the covariates and the theoretical medians of the lagged dependent latent variables.

Joint Distribution

For parameter estimation through a Gibbs sampler we need to derive the model joint distribution. Let $\mathbf{R}^*$ denote the $(T+1) \times 3$ matrix, the tth row of which is $\mathbf{r}_t^* = (r_t^{0*}, r_t^{-*}, r_t^{+*})$, $t = 0, 1, .., T$ and the three columns of which are $\mathbf{r}^{0*}$, $\mathbf{r}^{-*}$ and $\mathbf{r}^{+*}$; Δy denote the vector of dependent variables $(\Delta y_1, \ldots, \Delta y_T)$; θ denote the vector of model parameters $(\beta, \phi, \mathbf{c})$, where $\beta = (\beta^0, \beta^0, \beta^+)$, $\phi = (\phi^0, \phi^-, \phi^+)$, $\mathbf{c} = (\mathbf{c}^0, \mathbf{c}^-, \mathbf{c}^+)$, $\mathbf{c}^0 = (c_1^0, c_2^0)$, $\mathbf{c}^- = (c_{J^-}^-, \ldots, c_{-1}^-)$, and $\mathbf{c}^+ = (c_1^+, \ldots, c_{J^+}^+)$. The joint density for $\mathbf{R}^*$, Δy and θ can be factorized as

$$f(\Delta y, \mathbf{R}^*, \theta) = \left[\prod_{t=1}^{T} f(\Delta y_t, \mathbf{r}_t^* | \Delta y_1, \ldots, \Delta y_{t-1}, \mathbf{r}_0^*, \ldots, \mathbf{r}_{t-1}^*, \theta)\right]\pi(\mathbf{r}_0^*, \theta),$$

where π is the prior distribution. By the Markov property it reduces to

$$f(\Delta y, \mathbf{R}^*, \theta) = \left[\prod_{t=1}^{T} f(\Delta y_t, \mathbf{r}_t^* | \Delta y_{t-1}, \mathbf{r}_{t-1}^*, \theta)\right]\pi(\mathbf{r}_0^*, \theta).$$

Δy_t is fully determined by the latent variables $\boldsymbol{r}_t^*$ and cutpoints $\boldsymbol{c}$; and $\boldsymbol{r}_t^*$ in turn is determined by $\boldsymbol{r}_{t-1}^*$, $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$. Hence, the joint density factorizes further into

$$f(\boldsymbol{\Delta y}, \boldsymbol{R}^*, \boldsymbol{\theta}) = \left[\prod_{t=1}^T f(\Delta y_t | \boldsymbol{r}_t^*, \boldsymbol{c}) f(\boldsymbol{r}_t^* | \boldsymbol{r}_{t-1}^*, \boldsymbol{\beta}, \boldsymbol{\phi}) \right] \pi(\boldsymbol{r}_0^*, \boldsymbol{\theta}).$$

So we have to determine the factors $f(\Delta y_t | \boldsymbol{r}_t^*, \boldsymbol{c})$ and $f(\boldsymbol{r}_t^* | \boldsymbol{r}_{t-1}^*, \boldsymbol{\beta}, \boldsymbol{\phi})$. The first one is a sum of indicators of whether $\boldsymbol{r}_t^*$ lie between the corresponding cutpoints:

$$f(\Delta y_t | \boldsymbol{r}_t^*, \boldsymbol{c}) = I_{r_t^{0*} < c_1^0, r_t^{-*} \in [c_{\Delta y_t-1}^-, c_{\Delta y_t}^-]} + I_{r_t^{0*} \in [c_1^0, c_2^0], \Delta y_t = 0} + I_{r_t^{0*} > c_2^0, r_t^{+*} \in [c_{\Delta y_t}^+, c_{\Delta y_t+1}^+]}. \tag{4}$$

The probability density of $\boldsymbol{r}_t^*$ is given by

$$f(\boldsymbol{r}_t^* | \boldsymbol{r}_{t-1}^*, \boldsymbol{\beta}, \boldsymbol{\phi}) \propto \prod_{i=0,+,-} \exp \left[-\frac{1}{2} \left(r_t^{i*} - \boldsymbol{x}_t^{i'} \boldsymbol{\beta}^i - \phi^i r_{t-1}^{i*} \right)^2 \right].$$

Finally, we have to specify a prior π. We choose a uniform prior on the real numbers $\mathbb{R}$, i.e. an improper prior, for almost all parameters. The only restriction we have to consider is $\phi^i \in (-1, 1)$ to ensure stationarity of the latent AR [Eqs. (1) and (2)]. Besides, we choose the initial latent variables to be distributed as $r_0^{i*} \sim \mathcal{N}(0, (\tau)^2 = 10^2)$ for numerical reasons that will be apparent later. Finally, each pair of cutpoints is uniformly distributed only on the subset of $\mathbb{R}^2$, where the ordering $c_m^i < c_{m+1}^i$ is satisfied.

Challenges to embedding the model into a New Keynesian (NK) framework

Embedding the proposed SwAOP model into a NK DSGE framework would require substantial modifications and would cause serious computational issues.

The SwAOP model is designed for changes to the interest rates — it represents a policy rule in differences. The monetary policy equations in the NK DSGE models are formulated in terms of the levels of interest rates — typically they are represented by the Taylor-type rule. While the SwAOP model can be specified with the Taylor rule's variables (and its underlying equations for the latent continuous variables are linear, and can be interpreted as the optimal policy rules in differences, derived from minimizing the central bank's quadratic loss function), the NK model would need to be changed to work in changes, not the levels of the policy interest rate.[2]

The solution methods for the NK model would need to be adapted to allow for only discrete changes in the policy rates rather than continuous changes. The NK modeler would need to go far beyond the conventional log-linearization around a steady state and to use complicated statistical inference for the NK agents who do not observe the optimal interest rate. The technical problems with the computation of the equilibria, caused by the discreteness and nonlinearity of the policy rule, would become exacerbated by the unobserved switching among three latent regimes.

3. Estimation and inference

While the classical (not simulation-based) statistical approaches, such as maximum likelihood or method of moments, will not work for this model we can estimate the parameters and choice probabilities using a Gibbs sampler based on the MCMC algorithm for the AOP model with a grouped move step that drastically accelerates the convergence of the Gibbs sampler (Müller and Czado, 2005). The structure of the Gibbs sampler is as follows:

- Update for the latent variables: Sample the latent dependent variables $\boldsymbol{R}^*$ from the corresponding truncated normal distributions.
- Update for the cutpoint parameters: Sample the cutpoint parameters $\boldsymbol{c}$ from the uniform distributions.
- Update for the slope parameters: Sample the slope parameters $\boldsymbol{\beta}$ and the AR parameters $\boldsymbol{\phi}$ from the multivariate normal distributions.
- Grouped move step: Sample a scale parameter from a Gamma distribution to rescale the parameters $\boldsymbol{c}$, $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$ and the latent variables $\boldsymbol{R}^*$.

We use no assumption on the signs of the slope or AR parameters. Without loss of generality and to match the data on the changes to the target in our empirical application, we restrict our inference to the case of five outcome categories $j \in \{-2, -1, 0, 1, 2\}$ of the dependent variable Δy_t with two unknown cutpoints in each latent decision: $\boldsymbol{c}^0 = (c_1^0, c_2^0)$, $\boldsymbol{c}^- = (c_{-2}^-, c_{-1}^-)$ and $\boldsymbol{c}^+ = (c_1^+, c_2^+)$. The initial values for the Gibbs sampler for all three latent decisions $i \in \{0, -, +\}$ are $\phi_{\text{ini}}^i = 0$, $\boldsymbol{\beta}_{\text{ini}}^i = \boldsymbol{0}$ and $(c_{\text{l, ini}}^i, c_{\text{u, ini}}^i) = (-0.5, 0.5)$.

We will now derive all full conditional densities from the joint distribution exploiting the fact that each full conditional density is proportional to the joint distribution.

---

[2] Transforming the SwAOP model from a policy rule in differences to a policy rule in levels will lead to serious problems with estimation. On the left-hand side, instead of five discrete choices for the changes to the target ('large cut', 'small cut', 'no change', 'small hike' and 'large hike'), there will be forty discrete choices for the levels of the target (0%, 0.25%, ..., 9.5%, 9.75%). There will be small numbers of observations for many choices (often only 1–5 observations). We will not be able to reliably estimate the numerous parameters of the model.

3.1. Update for the latent variables

For convenience we use the abbreviation *rem* for all other remaining parameters and latent variables. For instance, the full conditional for the latent variable r_t^{0*} will be written as

$$f(r_t^{0*}|rem) := f(r_t^{0*}|\Delta y, \boldsymbol{r}^{+*}, \boldsymbol{r}^{-*}, r_0^{0*}, \ldots, r_{t-1}^{0*}, r_{t+1}^{0*}, \ldots, r_T^{0*}, \boldsymbol{\theta}).$$

The full conditional of the latent variable r_t^{i*}, $t \in \{1, 2, \ldots, T-1\}$, $i \in \{0, -, +\}$ is

$$f(r_t^{i*}|rem) \propto f(\Delta y_t|\boldsymbol{r}_t^*, \boldsymbol{c}) f(\boldsymbol{r}_t^*|\boldsymbol{r}_{t-1}^*, \boldsymbol{\beta}, \boldsymbol{\phi}) f(r_{t+1}^*|\boldsymbol{r}_t^*, \boldsymbol{\beta}, \boldsymbol{\phi})$$

$$\propto f(\Delta y_t|\boldsymbol{r}_t^*, \boldsymbol{c}) \exp\left[-\frac{1}{2}\left(r_t^{i*} - \boldsymbol{\beta}^i \boldsymbol{x}_t^i - \phi^i r_{t-1}^{i*}\right)^2 - \frac{1}{2}\left(r_{t+1}^{i*} - \boldsymbol{\beta}^i \boldsymbol{x}_{t+1}^i - \phi^i r_t^{i*}\right)^2\right]$$

$$\propto f(\Delta y_t|\boldsymbol{r}_t^*, \boldsymbol{c}) \exp\left[-\frac{1}{2}\left(\frac{r_t^{i*} - \mu_t^i}{\sigma^i}\right)^2\right],$$

where $(\sigma^i)^2 := [1 + (\phi^i)^2]^{-1}$ and $\mu_t^i := (\sigma^i)^2 [\boldsymbol{\beta}^i \boldsymbol{x}_t^i + \phi^i(r_{t-1}^{i*} + r_{t+1}^{i*} - \boldsymbol{\beta}^i \boldsymbol{x}_{t+1}^i)]$. Thus, r_t^{i*} must be sampled from a normal distribution $r_t^{i*}|rem \sim \mathcal{N}_{\text{trunc}_t^i}(\mu_t^i, (\sigma^i)^2)$, which is truncated by $f(\Delta y_t|\boldsymbol{r}_t^*, \boldsymbol{c})$ if the observation is influenced by r_t^{i*} (this is not always the case: e.g., if $\Delta y_t = 1$ or 2, r_t^{-*} can be arbitrary).

The truncation intervals for the regime decision r_t^{0*} are $\text{trunc}_t^0 = (a_t^0, b_t^0)$ with

$$a_t^0 = \begin{cases} c_2^0, & \text{if } \Delta y_t > 0, \\ c_1^0 & \text{if } \Delta y_t = 0, r_t^{-*} < c_{-1}^-, \\ -\infty & \text{else,} \end{cases} \quad b_t^0 = \begin{cases} c_1^0 & \text{if } \Delta y_t < 0, \\ c_2^0 & \text{if } \Delta y_t = 0, r_t^{+*} > c_1^+, \\ \infty & \text{else.} \end{cases}$$

The truncations for the amount decisions in the dovish and hawkish regimes are

$$\text{trunc}_t^- = \begin{cases} (c_{\Delta y_t - 1}^-, c_{\Delta y_t}^-) & \text{if } r_t^{0*} < c_1^{0*}, \\ (-\infty, \infty) & \text{else,} \end{cases} \quad \text{trunc}_t^+ = \begin{cases} (c_{\Delta y_t}^+, c_{\Delta y_t + 1}^+) & \text{if } r_t^{0*} > c_2^{0*}, \\ (-\infty, \infty) & \text{else.} \end{cases}$$

The updates for r_0^{i*} and r_T^{i*} slightly differ due to a missing predecessor or a successor. In addition, the density of r_0^{i*} is never truncated since there are no observations for it, and is given by

$$f(r_0^{i*}|rem) \propto f(\boldsymbol{r}_1^*|\boldsymbol{r}_0^*, \boldsymbol{\beta}, \boldsymbol{\phi}) \pi(\boldsymbol{r}_0^*)$$

$$\propto \exp\left[-\frac{1}{2}\left(r_1^{i*} - \boldsymbol{\beta}^i \boldsymbol{x}_1^i - \phi^i r_0^{i*}\right)^2\right] \exp\left[-\frac{1}{2}\left(\frac{r_0^{i*}}{\tau}\right)^2\right]$$

$$\propto \exp\left[-\frac{1}{2}\left(\frac{r_0^{i*} - \mu_0^i}{\sigma_0^i}\right)^2\right],$$

where $\mu_0^i := \phi^i(r_1^{i*} - \boldsymbol{\beta}^i \boldsymbol{x}_1^i)(\sigma_0^i)^2$ and $(\sigma_0^i)^2 := [(\phi^i)^2 + (\tau)^{-2}]^{-1}$, and the prior hyperparameter $\tau = 10$ is chosen to avoid a big variance $(\sigma_0^i)^2$ for parameters ϕ^i close to zero. In other words, r_0^{i*} must be sampled from $r_0^{i*}|rem \sim \mathcal{N}(\mu_0^i, (\sigma_0^i)^2)$.

Finally r_T^{i*} has the following full conditional density:

$$f(r_T^{i*}|rem) \propto f(\Delta y_T|\boldsymbol{r}_T^*, \boldsymbol{c}) f(\boldsymbol{r}_T^*|\boldsymbol{r}_{T-1}^*, \boldsymbol{\beta}, \boldsymbol{\phi})$$

$$\propto (\Delta y_T|\boldsymbol{r}_T^*, \boldsymbol{c}) \exp\left[-\frac{1}{2}\left(r_T^{i*} - \boldsymbol{\beta}^i \boldsymbol{x}_T^i - \phi^i r_{T-1}^{i*}\right)^2\right]$$

$$= (\Delta y_T|\boldsymbol{r}_T^*, \boldsymbol{c}) \exp\left[-\frac{1}{2}\left(r_T^{i*} - \mu_T^i\right)^2\right]$$

with mean $\mu_T^i := \boldsymbol{\beta}^i \boldsymbol{x}_T^i + \phi^i r_{T-1}^{i*}$; hence, r_T^{i*} must be sampled from $r_T^{i*}|rem \sim \mathcal{N}_{\text{trunc}_T^i}(\mu_T^i, 1)$. The interval trunc_T^i is the same as trunc_t^i, $t \in \{1, 2, \ldots, T-1\}$.

The sampling of the $3(T+1)$ latent variables $\boldsymbol{R}^*$ in this section constitutes the bottleneck of the whole Gibbs sampler. To speed up the sampler, every second step can be sampled simultaneously since r_t^* depends only on r_{t-1}^* and r_{t+1}^*.

3.2. Update for the slope parameters

We denote the parameters on all regressors in each latent decision $i \in \{0, -, +\}$ by $\boldsymbol{B}^i := (\boldsymbol{\beta}^i, \phi^i)$ and all regressors in each latent equation by a $T \times (k^i + 1)$ matrix $\boldsymbol{Z}^i := (\boldsymbol{X}^{i\prime}, \boldsymbol{r}_{-T}^{i*})$, the tth row of which is $\boldsymbol{z}_t^i := (\boldsymbol{x}_t^{i\prime}, r_{t-1}^{i*})$, $t \in \{1, 2, \ldots, T\}$, where $\boldsymbol{r}_{-T}^{i*} := (r_0^{i*}, r_1^{i*}, \ldots, r_{T-1}^{i*})'$. The full conditional density for $\boldsymbol{B}^i$ is

$$f(\mathbf{B}^i | rem) \propto \left[\prod_{t=1}^{T} f(r_t^{i*} | r_{t-1}^{i*}, \mathbf{B}^i) \right] \times \pi(\mathbf{B}^i)$$

$$\propto \exp\left\{ -\frac{1}{2} \left[\sum_{t=1}^{T} (r_t^{i*} - \boldsymbol{\beta}^i \mathbf{x}_t^i - \phi^i r_{t-1}^{i*})^2 \right] \right\} \times I_{|\phi^i|<1}$$

$$\propto \exp\left\{ 2 \left[\sum_{t=1}^{T} r_t^{i*} \mathbf{z}_t^{i'} \right] \mathbf{B}^i - \frac{1}{2} \mathbf{B}^{i'} [\sum_{t=1}^{T} \mathbf{z}_t^i \mathbf{z}_t^{i'}] \mathbf{B}^i \right\} \times I_{|\phi^i|<1}.$$

Note that $\mathbf{z}_t^i \mathbf{z}_t^{i'}$ is not a scalar but a dyadic product. So all $\mathbf{B}^i$ must be sampled independently from multivariate normal distributions

$$\mathbf{B}^i | rem \sim \mathcal{N}_{k^i+1}[(\mathbf{Z}^{i'} \mathbf{Z}^i)^{-1} \mathbf{Z}^{i'} \mathbf{r}_{-0}^{i*}, (\mathbf{Z}^{i'} \mathbf{Z}^i)^{-1}] \times I_{|\phi^i|<1},$$

which is truncated by $|\phi^i| < 1$ in one dimension, and where $\mathbf{r}_{-0}^{i*} := (r_1^{i*}, r_2^{i*}, \ldots, r_T^{i*})'$.

3.3. Update for the cutpoint parameters

The full conditional density $f(c_m^i | rem) \propto \prod_{t=1}^{T} f(\Delta y_t | \mathbf{r}_t^*, \mathbf{c}) \pi(\mathbf{c})$ for each cutpoint parameter in each latent decision i is uniform over an interval (l_m^i, u_m^i), i.e. $c_m^i | rem \sim \mathcal{U}(l_m^i, u_m^i)$, where the lower l_m^i and upper u_m^i bounds are determined by:

$$
\begin{array}{llll}
c_1^0 : & l_1^0 = \max\left\{ r_t^{0*} | \Delta y_t = -1, -2 \right\}, & u_1^0 = \min\left\{ c_2^0, r_t^{0*} | \Delta y_t = 0, r_t^{-*} < c_{-1}^- \right\}, \\
c_2^0 : & l_2^0 = \max\left\{ c_1^0, r_t^{0*} | \Delta y_t = 0, r_t^{+*} > c_1^+ \right\}, & u_2^0 = \min\left\{ r_t^{0*} | \Delta y_t = 1, 2 \right\}, \\
c_{-2}^- : & l_{-2}^- = \max\left\{ r_t^{-*} | \Delta y_t = -2 \right\}, & u_{-2}^- = \min\left\{ c_{-1}^-, r_t^{-*} | \Delta y_t = -1 \right\}, \\
c_{-1}^- : & l_{-1}^- = \max\left\{ c_{-2}^-, r_t^{-*} | \Delta y_t = -1 \right\}, & u_{-1}^- = \min\left\{ r_t^{-*} | \Delta y_t = 0, r_t^{0*} < c_1^0 \right\}, \\
c_1^+ : & l_1^+ = \max\left\{ r_t^{+*} | \Delta y_t = 0, r_t^{0*} > c_2^0 \right\}, & u_1^+ = \min\left\{ c_2^+, r_t^{+*} | \Delta y_t = 1 \right\}, \\
c_2^+ : & l_2^+ = \max\left\{ c_1^+, r_t^{+*} | \Delta y_t = 1 \right\}, & u_2^+ = \min\left\{ r_t^{+*} | \Delta y_t = 2 \right\}.
\end{array}
$$

These restrictions do not allow the cutpoint parameters to move substantially within each Gibbs step and slow down the convergence of the sampler.

3.4. Grouped move step

The problem with the slowly moving cutpoints was solved for the AOP model by using grouped move steps (Müller and Czado, 2005). To develop a suitable grouped move step for the SwAOP model, we employ a theorem from Liu and Sabatti (2000) that states:

If Γ is a locally compact group of transformations defined on the sample space S, L its left Haar measure with density l, $\boldsymbol{\omega} \in S$ follows a distribution with density f, and $\gamma \in \Gamma$ is drawn from the density $f_{\boldsymbol{\omega}}(\gamma) = f(\gamma(\boldsymbol{\omega}))|J_\gamma(\boldsymbol{\omega})|l(\gamma)$, where $|J|$ is the Jacobian determinant of this transformation. Then $\boldsymbol{\omega}^* := \gamma(\boldsymbol{\omega})$ has the density f too.[3]

To apply the theorem to our case (where the invariant density in the theorem f is the posterior) we write all latent variables and parameters into one vector $\boldsymbol{\omega} = (\mathbf{r}^{0*'}, \mathbf{r}^{-*'}, \mathbf{r}^{+*'}, \boldsymbol{\theta})$ and exploit the fact that the joint density is a product of normal densities $f(r_t^* | r_{t-1}^*, \boldsymbol{\beta}, \boldsymbol{\phi}) \propto \prod_{i=0,+,-} \exp\left[-\frac{1}{2}(\varepsilon_t^i)^2 \right] \times I$, where I is the product of the indicator functions in (4), which is $I = 1$ if $\mathbf{R}^*$, $\mathbf{c}$ and $\Delta\mathbf{y}$ match each other and $I = 0$ otherwise, and the error terms are $\varepsilon_t^i = r_t^{i*} - \boldsymbol{\beta}^i \mathbf{x}_t^i - \phi^i r_{t-1}^{i*}$. If we rescale the latent variables and the cutpoint parameters with the same value $g > 0$, the value of the indicator function I is unaffected (e.g., if $r_t^{0*} < c_1^0 \Leftrightarrow g r_t^{0*} < g c_1^0$). To get a simple distribution for the sampling of g we must also rescale the slope parameters to $g\boldsymbol{\beta}$. Therefore, we use the transformation

$$\gamma_g(\boldsymbol{\omega}) = (g\mathbf{r}^{0*'}, g\mathbf{r}^{-*'}, g\mathbf{r}^{+*'}, g\boldsymbol{\beta}, g\mathbf{c}, \boldsymbol{\phi}),$$

in which we rescale everything except for the AR parameters $\boldsymbol{\phi}$. The group Γ is, hence, the group of positive real numbers with the multiplication $\{\mathbb{R}_{>0}, \cdot\}$.

The left Haar measure of this group has the density $l(\gamma_g) = 1/g$ and Jacobian determinant $|J_\gamma(\boldsymbol{\omega})| = |\partial\gamma(\boldsymbol{\omega})/\partial\boldsymbol{\omega}| = g^p$, where $p = 3T + k^0 + k^- + k^+ + C$ is the number of scaled parameters ($C = 6$ is the number of cutpoint parameters). Then the density for the scaling parameter $g > 0$ is

$$f_g(g) = g^{p-1} \pi(\gamma_g(\boldsymbol{\omega})) \propto g^{p-1} \exp\left[-\frac{1}{2} g^2 \sum_i \sum_t (\varepsilon_t^i)^2 \right].$$

---

[3] A measure L on a group Γ is called a left Haar measure, if for all elements γ and all measureable subsets A of this group the measure is invariant under application of γ, i.e. $L(A) = L(\gamma A)$.

We abbreviate $q := \frac{1}{2} \sum_i \sum_t \left(\varepsilon_t^i \right)^2$ and compute the density of g^2 as

$$f_{g^2}(g^2) = f_g(g) \frac{\partial g}{\partial g^2} \propto (g^2)^{\frac{p}{2}-1} \exp \left[-qg^2 \right],$$

which is the density of a Gamma distribution $\Gamma(a, q)$ for g^2 with a shape parameter $a = \frac{p}{2}$ and a rate parameter q. The accelerating grouped move step consists of drawing a number g^2 from this distribution and multiplying the corresponding parameters with g. In fact, because the joint distribution factorizes into three parts from the three decisions $i \in \{0, -, +\}$, it is even possible to perform three independent grouped moves for the parameters of each decision separately. This will lead to an even higher variation and a faster convergence of the Gibbs sampler.

The efficiency of the grouped move can be explained as follows. Since the expectation and variance of the Gamma distribution are $\mathbb{E}_{\Gamma(a,q)}[g^2] = \frac{a}{q}$ and $\text{Var}_{\Gamma(a,q)}[g^2] = \frac{a}{q^2}$, both Gamma distribution parameters are huge numbers in our case. The summands of q are the squares of the error terms ϵ_t^i. Assuming both our model and the current Gibbs parameters are true, these increments are standard normally distributed and their square is χ^2 distributed with expectation value 1. Thus, close to the true parameters, we have $a \approx 3T/2 \approx \mathbb{E}[q]$. The expectation value for g is then $E_{\Gamma(a,q)}[g^2] \approx 1$ and the variance $\text{Var}_{\Gamma(a,q)}[g^2] = \frac{a}{q^2} \approx \frac{2}{3T}$ is small. This means that almost no rescaling happens (all sampled values for g will be close to 1) if the algorithm has almost converged. Far from convergence, e.g. $\boldsymbol{\omega} \approx (\lambda \boldsymbol{r}^{0*\prime}, \lambda \boldsymbol{r}^{-*\prime}, \lambda \boldsymbol{r}^{+*\prime}, \lambda \boldsymbol{\beta}, \lambda \boldsymbol{c}, \boldsymbol{\phi}) = \gamma_\lambda(\boldsymbol{\omega}_{\text{likely}})$, we have $\mathbb{E}[q] \approx 3\lambda^2 T/2 \Rightarrow E_{\Gamma(a,q)}[g^2] \approx \frac{1}{\lambda^2}$. Thus, the variance is still small and g close to $\mathbb{E}[g] \approx 1/\lambda$ will be sampled, so the next (scaled) parameters will be closer to the true parameters $\gamma(\boldsymbol{\omega}) = g\boldsymbol{\omega} \approx \boldsymbol{\omega}_{\text{likely}}$.

4. Finite sample performance

We performed the Monte Carlo experiments to assess the finite sample bias and uncertainty of the estimates of the parameters, choice probabilities and MEs of covariates on choice probabilities (and their asymptotic standard errors) in the proposed SwAOP estimator, and to compare the performance of the AOP and SwAOP estimators under each of the true DGPs.

The design and results of Monte Carlo experiments are reported in Appendix A. The results support the asymptotic consistency of the employed Bayesian estimator, which also performs well in small samples. The experiments suggest that the single-equation AOP model delivers asymptotically biased estimates of the probabilities if the DGP is subject to the heterogeneity of status quo outcomes. The SwAOP estimator behaves much better under the AOP true DGP than the AOP estimator under the data generated using the SwAOP model (although the AOP model is not nested in the SwAOP model). When the SwAOP model is fitted to data from the AOP model, as sample size grows, the accuracy of the SwAOP estimator improves. However, when the AOP model is fitted to data from the SwAOP model, as sample size grows, the accuracy of the AOP estimator deteriorates.

5. Empirical application

We apply the SwAOP model to explain the FOMC decisions on the target during the 7/1987 – 1/2006 period under Greenspan's chairmanship, and to predict the target decisions out-of-sample for the 3/2006 – 6/2019 period. We contrast the in- and out-of-sample performance of the SwAOP model with those of the AOP model and linear model estimated by ordinary least squares (OLS) using the same set of explanatory variables, and also with that of the Taylor-type rule.

5.1. Data and model specifications

The target, a principal measure of U.S. monetary policy during the entire sample (*de facto* since the fall of 1982; see Thornton, 2005), is set by the FOMC either at the eight prescheduled meetings per year or sometimes at the extra (un-scheduled) meetings. The FOMC decision-making cycle — one and a half month between the meetings — is long enough to update the macroeconomic data that are targeted by the Fed (such as inflation, output, unemployment, etc.) and arrive at monthly frequency, and to accumulate enough data that arrive weekly or daily (the Fed, which faces uncertainty and incurs the costs in the case of rate reversal, prefers to "wait and see" and to react to more accumulated economic data to minimize the risk of the reversals).

From 1993 through 2019, there were seven target changes in the intermeeting periods, all caused by extraordinary circumstances or crises. Among the 114 FOMC decisions from 11/1992 to 1/2006 there were only seven intermeeting decisions (among which there were only five target changes: in 1994, 1998, and 2001). From 2/2006 through 12/2019, there were only two intermeeting moves (both in 2008).

Prior to 1993, most target changes implemented two weeks and even later after a scheduled meeting were unambiguously settled on that last meeting with the delay of implementation. As a matter of fact, prior to 1993 more than two thirds of target changes are executed between the scheduled FOMC meetings. Out of the 76 FOMC decisions from 7/1987 to 10/1992, 33 decisions (among which there were 29 changes) were made at the unscheduled meetings and teleconferences, and 43 decisions (among which there were only 12 changes) were made at the scheduled meetings.

The Fed changed its operating procedure after 1992, and not only stopped making hidden policy decisions at the scheduled meetings (to be implemented one, two or three weeks later) but also drastically reduced (and eventually abandoned)

its practice to run unscheduled meetings without extraordinary circumstances. After 1992, if the change was desirable in the intermeeting period, the FOMC typically made no changes in between the meetings and waited until the next scheduled meeting (unless a rear extraordinary event called for an immediate reaction). The observed policy inactions in the intermeeting periods may not reflect the actual Fed's response to the economic developments, but rather the Fed's desire to obtain more accumulated data and reluctance to change the target between the scheduled meetings.

To model the target, instead of using the quarterly or monthly averages of the federal funds rate and economic variables as is common practice in the literature, we employ the data at the frequency of FOMC decisions, both scheduled and unscheduled ones. We use the historical dates of all FOMC decisions (made either at scheduled or unscheduled meetings, or occasionally at the discretion of the chairman during intermeeting periods) as sample observations with the following advantages.

First, we avoid a problem of reverse causation typical for time-aggregated data. Second, we avoid noise from periods with no movements in the target, when the observed policy inactions may not reflect the actual Fed response to economic developments, but rather the Fed reluctance to change the target between scheduled meetings, especially in the weeks prior to them. Third, information about the exact timing of FOMC meetings together with the use of daily financial data allows us to substantially improve the identification of the Fed policy rule. We match the FOMC decisions with the latest real-time vintages of monthly macroeconomic and non-aggregated daily financial data truly available before each meeting and not revised later on.

Omitting the unscheduled meetings (hence, ignoring the important information on the FOMC behavior), or interpreting the policy inactions between the scheduled meetings as the implicit decisions to not adjust the rate (and perhaps even augmenting the sample by including these hypothetical intermeeting no-change decisions), or, moreover, using the monthly, weekly or daily frequency to model the FOMC decisions would distort the statistical inference about the Fed's policy responses.

The dates of FOMC decisions and the original (unconsolidated) target changes are reported in Table in the Online Appendix B, and are based on information available at ALFRED[4]. The observed target decisions have the following distribution: -50 bp – 13 times, -31.25 bp – one, -25 bp – 30, -18.75 bp – one, -12.5 bp – one, 0 bp – 99, 6.25 bp – three, 12.5 bp – two, 18.75 bp – three, 25 bp – 27, 31.25 bp – two, 43.75 bp – two, 50 bp – five, and 75 bp – one time. Since October 1989 the Fed made changes to the target only in increments of 25 bp. We classify target changes into five categories of the dependent variable Δy_t: a 'large cut' – a decrease 31.25 bp or more, a 'small cut' – a decrease 25 bp at most and 12.5 bp at least, 'no change' – either no change or a change no more than 6.25 bp, a 'small hike' – an increase 25 bp at most and 12.5 bp at least, and a 'large hike' – an increase 31.25 bp or more. The sample consists of 190 observations with 14, 32, 102, 32 and 10 observations in the above categories, respectively.[5]

The relationship between economic developments and the Fed response to them is often modeled using a simple policy rule such as the Taylor rule (Taylor, 1993), which establishes a simple linear relation between the policy interest rate, inflation and the output gap:

$$y_t^{eff} = r^* + \pi_t + 0.5 gap_t + 0.5(\pi_t - \pi^*),$$

where y_t^{eff} – the effective federal funds rate, r^* – the equilibrium real interest rate, π_t – the inflation rate over the previous four quarters, π^* – the long-run inflation target, and gap_t – the output gap (the percent deviation of the real gross domestic product (GDP) from the potential one). Unfortunately, the equilibrium rate and potential output are not observed. To provide a better benchmark for the target decisions we estimate a more general (with interest-rate smoothing) version of the Taylor rule using OLS and publicly available real-time data:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 \pi_t + \alpha_3 gap_t,$$

where π_t – the real-time forecast of the core inflation rate for the current quarter, seasonally adjusted, annualized percentage points (before 1/2000 — the Greenbook projection for quarter-over-quarter core consumer price index inflation rate; from 1/2000 through 12/2013 — the Greenbook projection for quarter-over-quarter core personal consumption expenditures (PCE) inflation rate, chain weight; since 1/2014 — the median forecast of the annualized quarter-over-quarter percent changes of the core PCE inflation rate for the current quarter from the Survey of Professional Forecasters (SPF) provided by Philadelphia Fed); gap_t – the real-time forecast of the output gap, the difference between the actual and potential output expressed as a percent of potential output, for the current quarter (before 1/2014 — the Greenbook projections; since 1/2014 — the projections derived from the Congressional Budget Office's estimates of potential real GDP and SPF's median forecasts of the real GDP, seasonally adjusted); and $\alpha_0, \alpha_1, \alpha_2$ and α_3 — the unknown OLS parameters to be estimated. The real-time data on y_t, y_{t-1}, π_t and gap_t are retrieved from Philadelphia Fed's RTDSM[6] and ALFRED. The nowcasts of π_t and gap_t for the current quarter provided similar or better explanatory power than the one-, two-, three- and four-quarter ahead forecasts.

---

[4] ALFRED (ArchivaL Federal Reserve Economic Data) is available at https://alfred.stlouisfed.org/

[5] A 'symmetrical' classification, namely: [-50 bp, -37.5 bp] is 'a large cut', (-37.5 bp, -12.5 bp] is 'a small cut', (-12.5 bp, 12.5 bp) is 'no change', [12.5 bp, 37.5 bp) is a 'small hike', and [37.5 bp, 75 bp] is a 'large hike' produces a similar distribution of the dependent variable (13, 33, 102, 34, 8) and similar estimation results.

[6] RTDSM (Real-Time Data Set for Macroeconomists) is available at https://www.philadelphiafed.org.

Recent studies, using discrete-choice models for the target, document that financial indicators, such as the spread between the long- and short-term interest rates, explain the Fed policy decisions better than the Taylor-rule variables (Hamilton and Jorda, 2002; Kauppi, 2012; Piazzesi, 2005). The FOMC always starts its meetings with a review of the 'financial outlook'. The spread can be interpreted as a market-based proxy of future inflation and real activity (Estrella and Hardouvelis, 1991; Frankel and Lown, 1994; Mishkin, 1990). In addition to the spread, we employ two forward-looking indicators of the economic situation computed by Fed staff in the Greenbook for each FOMC meeting: the forecast of housing starts (one of the leading indicators of economic activity, frequently mentioned in the FOMC documents) and the forecast of GDP growth (the economic growth is one of the four goals of the Fed's monetary policy). Finally, we construct a Fed-based proxy of future inflation and real activity, derived from the FOMC documents. Together with its decision on the target, the FOMC issued (before 2000) a statement about its expectations of the future stance of monetary policy and (since 2000) a statement on the balance of risks for inflation and economic growth. These post-meeting statements have been widely seen as the indicators of the next FOMC actions, and are shown to contain predictive content for forecasting target even after controlling for macroeconomic variables (Hayo and Neuenkirch, 2010; Lapp and Pearce, 2000; Pakko, 2005). The FOMC statements and interest rate spreads summarize the vast amount of forward-looking information available both to the Fed in setting the policy rate and to Fed watchers in anticipating the Fed decisions, and can help surmounting the omitted variables and time-varying parameter problems (Cochrane and Piazzesi, 2002).

Thus, the explanatory variables in the linear, AOP and SwAOP models include:

(i) $spread_t$ — the difference between the one-year treasury constant maturity rate and the effective federal funds rate, three-business-day moving average (the real-time data source: ALFRED);

(ii) $houstart_t$ — the nowcast of the total number of new privately owned housing units started (in thousands) for current quarter[7] (before 1/2014 — the Greenbook projections; since 1/2014 — the latest available monthly data on housing starts[8]; the real-time data source: ALFRED and RTDSM);

(iii) gdp_t — the nowcast of the quarter-over-quarter growth in the nominal GDP for the current quarter[9], percent change at annual rate, seasonally adjusted (before 1/1992 — the Greenbook projections for the gross national product; from 1/1992 through 12/2013 — the Greenbook projections for the GDP; since 1/2014 — the New York Fed staff nowcast of the GDP; the real-time data source: RTDSM and New York Fed[10]).

(iv) $tight_{t-1}$ and (v) $easing_{t-1}$ — the two binary indicators constructed from the 'policy bias' or 'balance-of-risks' statements at the previous FOMC decision announcement: $tight_{t-1}$ is equal to one if the statement at the previous FOMC meeting was tightening, and zero otherwise; and $easing_{t-1}$ is equal to one if the statement was easing, and zero otherwise (the real-time data source: the FOMC statements and minutes[11]). The examples of the easing, neutral and tightening FOMC statements are given below.

An example of the easing 'balance-of-risks' statement from the FOMC press release on June 25, 2003:

> The Committee perceives that the upside and downside risks to the attainment of sustainable growth for the next few quarters are roughly equal. In contrast, the probability, though minor, of an unwelcome substantial fall in inflation exceeds that of a pickup in inflation from its already low level. On balance, the Committee believes that the latter concern is likely to predominate for the foreseeable future.

An example of the tightening 'balance-of-risks' statement from the FOMC press release on March 21, 2000:

> Against the background of its long-run goals of price stability and sustainable economic growth and of the information currently available, the Committee believes the risks are weighted mainly toward conditions that may generate heightened inflation pressures in the foreseeable future.

And an example of the neutral 'balance-of-risks' statement released on May 3, 2005:

> The Committee perceives that, with appropriate monetary policy action, the upside and downside risks to the attainment of both sustainable growth and price stability should be kept roughly equal.

The values of dependent and explanatory variables are reported in Table B1 in the Online Appendix B. Sample descriptive statistics are shown in Table B2 in the Online Appendix B. The covariates' first-order autocorrelation coefficients are between 0.63 and 0.98. According to the augmented Dickey-Fuller unit root test, the null hypothesis of a unit root is rejected for all employed variables but one at the 0.0001 significance level (see Table B3 in the Online Appendix B); $houstart_t$ seems to be stationary only at the 0.06 level. However, if we test the actual housing starts series, which is highly correlated with the

---

[7] The nowcast for the current quarter provides more explanatory power than for one, two, three and four quarters ahead.

[8] The Greenbook projections are not available in the real time, and available to public only with the five-year lag. The actual values of housing starts are highly correlated with the Greenbook projections for the current quarter (the first-order autocorrelation coefficient is 0.98), and yields similar results if applied to the entire sample.

[9] The nowcast for the current quarter provides more explanatory power than for one, two, three and four quarters ahead.

[10] https://www.newyorkfed.org/research/policy/nowcast.html

[11] https://www.federalreserve.gov/monetarypolicy/fomc_historical.htm.

Greenbook projections (the correlation coefficient is 0.98), but available for a far longer period at monthly frequency, we reject the null of a unit root at the 0.01 level.

5.2. Estimation results

Table 1 shows the estimation results for the Taylor rule, linear, AOP and SwAOP models. The slope parameters in the SwAOP model (see the right three columns in Table 1) have the expected signs and are statistically different from zero at the 0.05 significance level (according to the empirical confidence intervals from the Gibbs sampling and assuming the asymptotic normality of the posterior distribution). The latent continuous dependent variable r_t^{0*}, representing the degree of the Fed policy stance and determining the regime decision, is driven by $tight_{t-1}$ (if the 'policy bias' was tightening at the previous meeting, the probability of a tight regime at the next meeting is larger), $easing_{t-1}$ (if the 'policy bias' was easing, the probability of a tight regime is smaller), $spread_t$, $houstart_t$, gdp_t (the larger the covariate values, the larger the probability of a tight regime, and the smaller the probability of an easing regime), and its lagged value r_{t-1}^{0*} (the higher the degree of the policy stance at the previous meeting, the larger the probability of a tight regime at the next meting). The continuous latent dependent variables r_t^{-*} and r_t^{+*}, representing the desired amount of the target change in the easing and tight regimes, respectively, are driven by $spread_t$ (only in the easing regime), gdp_t (the larger the covariate value, the larger the probability of a higher target level), and their lagged values r_{t-1}^{-*} and r_{t-1}^{+*}, respectively (the larger their values, the smaller the probability of a higher target level). The coefficient on $spread_t$ is not significant in the tight regime. The coefficients on $houstart_t$, $tight_{t-1}$ and $easing_{t-1}$ are not significant in both amount equations.

The responses to the easing and tightening policy statements $tight_{t-1}$ and $easing_{t-1}$ are asymmetrical: the Fed seems to be much more eager to cut the target rate under the easing policy directive rather than to hike it under the tightening directive. The policy reactions to the economic growth and expected inflation (proxied by the interest rate spread) differ in the two regimes: in the tight regime (when the Fed is considering hikes to the target) it reacts stronger to gdp_t while in the easing regime (when the Fed is considering cuts) it pays more attention to $spread_t$.

An interesting finding is the opposite sign of the coefficients on the lagged dependent variables: the sign is positive in the regime equation (the *p*-value is 0.0019), but negative in the amount equations (the *p*-value is 0.0000 in the tight regime and 0.1328 in the easing regime). It implies that the regime and amount decisions have different dynamics. The positive autocorrelation in the regime equation leads to the persistency of regime decisions, whereas the negative autocorrelations in the amount equations mean that the larger the desired change (a cut or a hike) at the previous meeting, the more likely is a status quo decision at the next meeting. The Fed seems to deliberately smooth the path of its target — it prefers to wait and see and to avoid making consecutive changes. Such an inference is impossible if we estimate a single-equation AOP or a linear model with the same set of covariates as in the SwAOP model (see the third or second column in Table 1): the AR coefficients on the lagged dependent variable in the linear and the AOP models are small and not significant (at the 0.06 level in the linear model, and at the 0.51 level in the AOP model), implying a lack of intentional interest-rate smoothing by the Fed.

The estimated probabilities of the three policy regimes are shown in Fig. 3 for each FOMC decision during the Greenspan era. The sample average probabilities of easing, neutral and tight regimes are 0.47, 0.28 and 0.25, respectively, though the frequencies of observed cuts, status quo decisions and hikes are 0.24, 0.54 and 0.22, respectively. Apparently, not all status quo decisions are generated by the neutral policy stance. The decomposition of the probability of no change $\Pr(\Delta y_t = 0)$ into three components $\Pr(\Delta y_t = 0 | s_t = -1)$, $\Pr(\Delta y_t = 0 | s_t = 0)$ and $\Pr(\Delta y_t = 0 | s_t = 1)$ conditional on the easing, neutral and tight regimes is on average 0.42, 0.54 and 0.05, respectively. The amount decisions tend to smooth the target and moderate or offset the vast majority of the easing and tight policy stances: almost a half of the status quo decisions are generated in either easing or tight regimes.
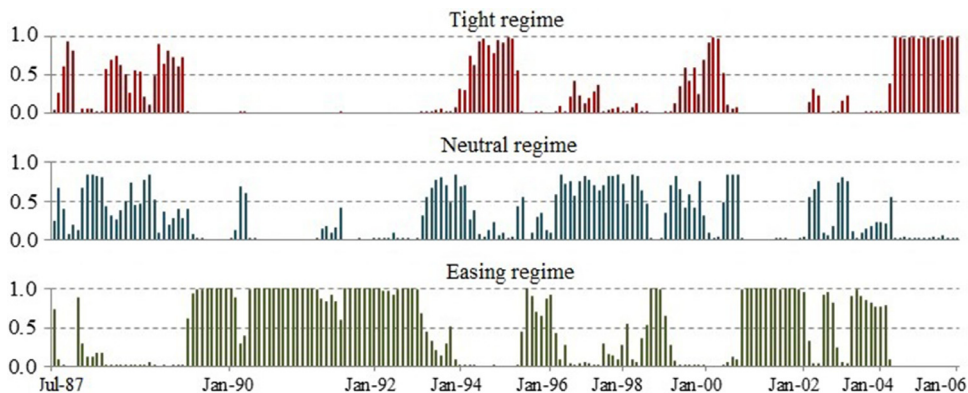

Fig. 3. In-sample probabilities of latent policy regimes at each FOMC meeting. Notes. Sample period: 7/1987–1/2006 (190 observations). The estimates are obtained from the SwAOP model (see Table 1).

Table 1

Parameter estimates of the Taylor rule, the linear, AOP and SwAOP models.

Variables	Taylor rule (OLS)	Linear model (OLS)	AOP model	SwAOP model		
				Regime equation	Amount equations	
					Easing regime	Tight regime
			Slope parameters			
tight		0.12(0.03)	0.66(0.24)	1.02(0.43)		
easing		-0.06(0.03)	-0.56(0.25)	-7.84(3.17)		
spread.		0.22(0.03)	1.39(0.21)	2.51(0.60)	1.19(0.32)	
houstart.		0.22(0.05)	1.44(0.39)	5.22(1.24)		
gdp		0.02(0.01)	0.29(0.07)	0.30(0.13)	0.34(0.11)	1.17(0.53)
π	0.11(0.03)					
gap	0.08(0.01)					
			Autoregressive parameters			
	0.92(0.02)	0.01(0.01)	0.05(0.08)	0.30(0.10)	-0.23(0.15)	-0.85(0.09)
	Intercept parameters			*Cutpoint parameters*		
	0.10(0.04)	-0.53(0.09)	1.21(0.71)	12.65(3.53)	-0.40*(0.32)	0.53(1.25)
			2.51(0.71)	15.90(3.68)	1.00(0.34)	5.69(2.14)
			5.36(0.81)			
			6.68(0.86)			

Notes. Sample period: 7/1987–1/2006 (190 observations). Standard deviations of parameters are in parentheses. Each FOMC decision in the sample is matched by the real-time values of covariates as they were known at the end of the previous day.

5.3. In- and out-of-sample comparison of competing models

The Taylor rule and three competing models, estimated with the same set of covariates, are contrasted in Table 2. We compare the in-sample fit for the Greenspan era and the out-of-sample one-step-ahead forecasting performance with re-cursive re-estimation with an increasing window for the next 111 FOMC decisions during the 3/2006–6/2019 period. It is a rather challenging forecasting exercise since during the in-sample period the target has always been far above zero, whereas out of the sample it abruptly moved forward zero, stuck close to zero for seven years, and then began gradually moving up (the 1/2016–6/2019 period). We compare the accuracy (the percentage of correct predictions), mean absolute error (MAE) and two strictly proper scores: the probability, or Brier score (Brier, 1950) and ranked probability score (Epstein, 1969). Both scores measure the accuracy of the probabilistic forecasts (contrary to the percentage of correct predictions that does not distinguish between cases in which the estimated probability of a particular choice is, for example, 0.51 or 0.99). Both scores have a minimum value of zero when all the observed choices are forecasted with a unit probability. In contrast to the Brier score, the ranked probability score punishes forecasts more severely for non-zero predicted probabilities of choices that are further from the observed choice. Probabilities for the AOP and SwAOP models are evaluated numerically as $\Pr(\Delta y_t = j | \boldsymbol{\Omega}) = \int \Pr(\Delta y_t = j | \boldsymbol{\Omega}, \boldsymbol{R}^*, \boldsymbol{\theta}) d \Pr(\boldsymbol{R}^*, \boldsymbol{\theta} | \boldsymbol{\Omega})$, where $\boldsymbol{\Omega}$ is the available data.

The SwAOP model outperforms the competitors both in and out of sample according to all employed criteria, whereas the Taylor rule demonstrate the worst results. The MAE in the SwAOP model is lower than in the AOP model by one third both in and out of sample. The performance of the linear model is better than that of the AOP model in sample but much worse out of sample. The out-of-sample forecasts of the Taylor rule and linear model are remarkably inferior to that of the discrete-choice competitors: the MAE is bigger by, respectively, 69% and 52%, and fraction of correct predictions is lower by 32 and 20 percentage points than those in the SwAOP model. In particular, during the ZLB period from 1/2009 through 12/2015, the SwAOP successfully accommodates the ZLB by a protracted switch to an easing regime (see Fig. 4), correctly predicts 53 out of 55 status quo decisions, and wrongly predicts only two further cuts. By contrast, the Taylor rule and linear model fail to handle the ZLB and correctly predict only 34 no-change decisions, wrongly predicting 21 further cuts.[12]

---

[12] The unconventional monetary policy tools such as large-scale asset purchases (commonly known as quantitative easing) and forward guidance took the central role in the conduct of policy during the ZLB period, not the target, lowering which further was not an option. Modeling the latent optimal shadow (continuous-valued, and possibly negative-valued) level of the policy interest rate as a new measure of monetary policy stance is the approach that can provide the further insights into the FOMC behavior during the ZLB period, and can allow to study the effects of unconventional monetary policy (see Wu and Xia, 2016, and references therein).

Table 2

Performance of competing models: the FOMC decisions on federal funds rate target favor the SwAOP model.

Model:	Taylor rule (OLS)	Linear (OLS)	AOP	SwAOP
In-sample forecast (7/1987 - 1/2006 period, 190 observations)				
Fraction of correct predictions	0.53	0.69	0.62	**0.74**
Mean absolute error, bp	14.0	7.9	10.7	**6.8**
Brier probability score			0.49	**0.37**
Ranked probability score			0.29	**0.21**
Out-of-sample forecast (3/2006 - 6/2019 period, 111 observations)				
Fraction of correct predictions	0.55	0.68	0.80	**0.87**
Mean absolute error, bp	13.7	9.0	6.8	**4.3**
Brier probability score			0.32	**0.25**
Ranked probability score			0.20	**0.16**

Notes. Specifications of all models are those reported in Table 1. Each FOMC decision is matched by the real-time covariates as they were known at the end of the previous day. The out-of-sample forecast of the next FOMC decision is performed using a recursive re-estimation with an increasing window. Predicted choice is that with the highest predicted probability (for the AOP and SwAOP models). Predicted choice in the Taylor rule and linear model is determined by rounding the continuous-valued prediction to the nearest discrete-valued choice. Brier probability score is computed as $\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{5}[\Pr(\Delta y_t = j) - d_{jt}]^2$, where indicator $d_{jt} = 1$ if $\Delta y_t = j$ and $d_{jt} = 0$ otherwise. Ranked probability score is computed as $\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{5}[P_{jt} - D_{jt}]^2$, where $P_{jt} = \sum_{i=1}^{j}\Pr(\Delta y_t = i)$ and $D_{jt} = \sum_{i=1}^{j}d_{it}$. The better the prediction, the smaller both score values. Mean absolute error is the average absolute difference between the observed and predicted discrete choice.
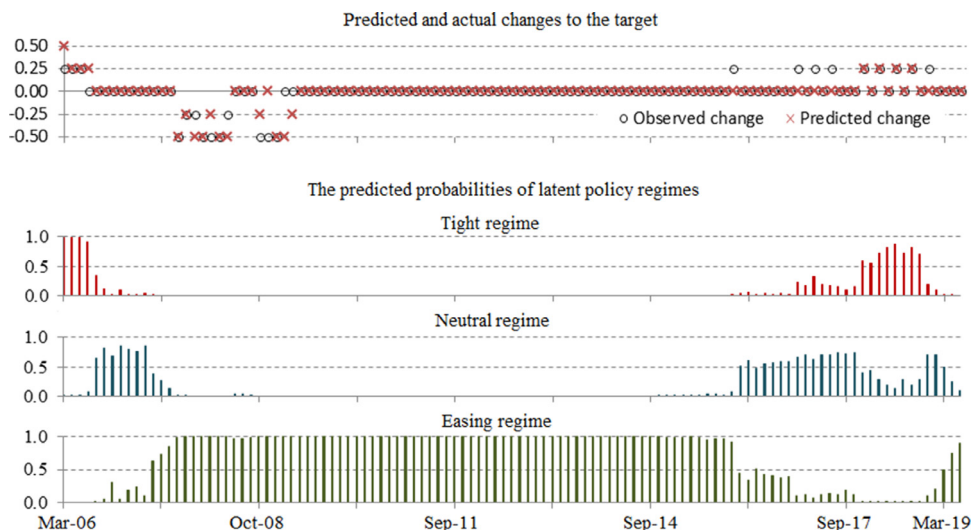
Fig. 4. Out-of-sample predicted changes to the target and probabilities of latent policy regimes at each FOMC meeting. Notes. Out-of-sample period: 3/2006–6/2019 (111 observations). The forecasts are obtained from the SwAOP model (see Table 1) using a recursive re-estimation with an increasing window.

The fractions of correct predictions in the SwAOP model in terms of three choices (hike, no change, cut), which are 0.79 in sample and 0.92 out of sample, exceed the accuracy of the existing dynamic discrete-choice models for the FOMC decisions such as those in Hu and Phillips (2004), Piazzesi (2005) and Kim et al. (2009). They model FOMC decisions made only at the scheduled meetings (which are easier to predict than unscheduled) in the 2/1994–12/2001, 2/1994–12/1998 and 2/1994–12/2001 periods, respectively, and report the following fractions of correct in-sample predictions: 0.78, 0.75 and 0.67. The out-of-sample forecast is performed only by Kim et al. (2009) for the 2002–2006 period with the 0.75 accuracy.

6. Concluding remarks

We develop a new dynamic discrete-choice model for monetary policy interest rates. Central banks typically adjust policy rates by discrete increments and often leave them unchanged in different economic circumstances; if central banks make a change, it is usually followed by further changes in the same direction. To address these stylized facts the new regime switching autoregressive ordered probit allows (i) the status quo outcomes to be heterogeneous and generated in three latent regimes, which can be interpreted as monetary policy stances (easing, neutral and tightening); (ii) the probabilities of positive and negative changes to the rate to be driven by different processes; and (iii) the persistence of policy rates and monetary policy inertia to be captured by the lagged dependent latent variables among the regressors.

The simulations demonstrate that the proposed Bayesian estimator performs well in small samples. The application to the FOMC decisions on the federal funds rate target shows that the discrete-choice approach and regime switching do matter in the empirical estimation of monetary policy rules. The proposed model overwhelmingly outperforms the linear models (including the Taylor rule), the single-equation ordered probit and other discrete-choice models both in and out of sample.

The new methodology can be employed to model the policy rate decisions of many central banks, changes to rankings, tick-by-tick stock price changes, and other ordinal data, which are subject to regime switching or sample separation and have abundant and heterogeneous observations in a middle neutral category (typically, a zero or no-change).

Appendix A. Finite sample performance

Monte Carlo design

The dependent variable is generated with five ordered choices. The values of parameters are calibrated to yield on average the following frequencies of the observed choices: 7%, 14%, 58%, 14% and 7%. The number of replications is 10,000 in each experiment. Three vectors of covariates v_1, v_2 and v_3 are drawn at each replication as $v_{1,t} \sim 0.41 v_{1,t-1} + \mathcal{N}(0, 0.3^2)$, $v_{2,t} \sim 0.17 v_{2,t-1} + \mathcal{N}(0, 0.4^2)$ and $v_{3,t} \sim 0.01 + 0.87 v_{3,t-1} + \mathcal{N}(0, 0.8^2)$. Since the dependent variable represents changes to policy interest rates, the simulated artificial covariates v_1, v_2 and v_3 mimic macroeconomic variables that are of interest to central banks such as changes to the inflation rate (personal consumption expenditures: chain-type price index, percent change from year ago), changes to the output gap, and the interest rate spread (one-year treasury constant maturity minus federal funds rate), respectively. The vectors of error terms in the latent Eqs. (1) and (2) are repeatedly generated as IID standard normal random variables.

Two competing models are simulated: the AOP and the SwAOP. For each true DGP, the repeated samples with 200, 500, and 1000 observations are generated as follows: (i) for the AOP DGP — with the covariates v_1, v_2 and v_3, the vector of slope parameters $(0.6, 0.8, 1)$, the vector of cut-point parameters $(-4.61, -2.45, 2.77, 4.95)$, and an AR coefficient 0.5; (ii) and for the SwAOP DGP— with $X = (v_1, v_2)'$, $X^- = X^+ = v_3$, $\boldsymbol{\beta} = (0.6, 0.8)$, $\boldsymbol{\beta}^- = 1.5$, $\boldsymbol{\beta}^+ = 2.4$, $c^0 = (-0.31, 0.31)$, $c^- = (-2.8, 0.31)$, $c^+ = (-0.02, 4.79)$, $\phi^0 = 0.5$, $\phi^- = 0.2$, and $\phi^+ = 0.2$.

Under each true DGP, the competing models are estimated using the same set of covariates: (i) under the AOP DGP— the AOP model is estimated with the covariates v_1, v_2 and v_3, and SwAOP model is estimated with $X = X^- = X^+ = (v_1, v_2, v_3)'$; (ii) and under the SwAOP DGP— the AOP model is estimated with v_1, v_2 and v_3, and the SwAOP model is estimated with $X = (v_1, v_2)'$ and $X^- = X^+ = v_3$. The simulations in (ii) are expected to provide some evidence that the AOP estimator deliver asymptotically biased estimates under the SwAOP DGP, whereas the simulations in (i) are expected to show that the SwAOP estimator provides asymptotically unbiased estimates under its own DGP and also even if the true DGP is the AOP model.

For each repeated estimation, we initialize the parameters at the starting point described in Section 3. We run the Gibbs sampler for a burn-in period with 80,000 iterations. We found that the Gibbs sampler converges much faster in most runs; if not, we produce 30,000 more iterations to approximate the posterior distribution. From this estimation period we compute the posterior means and all other estimates.

Monte Carlo results

First, we study the finite sample performance of the SwAOP estimator, when the data are generated by its own DGP with 200, 500 and 1000 observations. As Table A1 shows, the accuracy of the estimates improves with the number of observations. The small true values of ϕ^- and ϕ^+ imply a small positive AR effect, which, however, is detected even for only 200 observations. The root mean square errors (RMSE) of the parameters shrinks roughly with the root of the sample size. The biases do so as well, with an exception for a few cutpoint parameters. The results support the asymptotic consistency of the employed Gibbs sampler.

Next, we compare the performance of the AOP and SwAOP models under each DGP. The parameter estimates are not compatible due to the different structure of the two models. To compare them we also estimate the choice probabilities, the MEs of each covariate on each choice probability (the matrix of the MEs has $3 \times 5 = 15$ elements; their values, which depend on the values of the covariates, are computed at the population medians of the covariates and lagged latent dependent variables), and the standard errors of ME estimates. Table A2 shows the measures of accuracy of the ME estimates in both models under two alternative DGPs. If the simulated and estimated models are identical, the empirical coverage probabilities

Table A1
Monte Carlo results: the accuracy of parameter estimates in the SwAOP model.

Parameter	True value	Sample size:	Bias 200	500	1000	RMSE 200	500	1000
c_1^0	-0.31		0.005	0.016	0.019	0.226	0.148	0.108
c_2^0	0.31		-0.022	-0.015	-0.014	0.212	0.141	0.101
c_{-2}^-	-2.80		-0.774	-0.264	-0.106	1.193	0.578	0.324
c_{-1}^-	0.31		0.274	0.090	0.029	0.647	0.344	0.214
c_1^+	-0.02		0.024	-0.032	-0.022	0.549	0.327	0.227
c_2^+	4.79		0.914	0.564	0.311	1.321	0.974	0.699
ϕ^0	0.50		-0.012	-0.005	-0.002	0.097	0.059	0.042
ϕ^-	0.20		-0.027	-0.010	-0.005	0.136	0.083	0.058
ϕ^+	0.20		-0.022	-0.008	-0.003	0.110	0.067	0.046
β_1^0	0.60		0.046	0.015	0.009	0.323	0.190	0.133
β_2^0	0.80		0.062	0.020	0.011	0.287	0.167	0.116
β_3^-	1.50		0.563	0.198	0.087	0.836	0.377	0.208
β_3^+	2.40		0.624	0.342	0.186	0.897	0.565	0.388

Table A2
Monte Carlo results: the accuracy of ME estimates in the AOP and SwAOP models under each DGP.

Sample size	Simulated model: Estimated model:		AOP AOP	SwAOP	SwAOP AOP	SwAOP
200	Bias (the absolute difference between the estimated and true values of MEs)		0.03	0.12	5.41	0.47
500			0.01	0.06	6.07	0.22
1000			0.00	0.04	6.37	0.17
200	RMSE (the root mean square error of ME estimates relative to their true values)		3.19	4.09	6.74	7.76
500			1.59	2.18	6.71	5.31
1000			0.92	1.39	6.73	3.79
200	Coverage probability (the percentage of times the estimated 95% credibility intervals cover the true values of MEs)		90.1	68.6	34.4	92.4
500			93.0	73.0	18.1	93.5
1000			93.9	75.7	10.5	94.1
200	Standard deviation bias (the difference between the average of standard deviation estimates and the standard deviation of ME estimates)		2.02	1.79	0.61	2.28
500			1.04	1.01	0.44	1.88
1000			0.59	0.62	0.31	1.41

Notes. All reported measures of accuracy are averaged across all five choices and all three covariates. Bias, RMSE and standard deviation bias are multiplied by 100.

are quite close to 95% nominal level even for 200 observations and are getting even closer as sample size increases. The bias and RMSE are larger for the SwAOP model since it has more parameters than the AOP model. The biases and RMSE of both estimators seem to converge to zero. The biases of the standard deviation estimates also shrink with growing number of observations, which means the standard deviation of all posterior means is close to the average posterior standard deviation, fostering the reliability of the posterior.

The SwAOP estimator behaves much better under the AOP true DGP than the AOP estimator under the data generated using the SwAOP model. When the SwAOP model is fitted to data from the AOP model, as sample size grows, the biases

and RMSE decrease sharply, the coverage probability slowly approaches the nominal level and reaches 75% with 1000 observations. However, when the AOP model is fitted to data from the SwAOP model, as sample size grows, the bias increases, the RMSE remains almost the same, and the coverage probability deteriorates further from an already very bad value of 34% with 200 observations toward about 10% with 1000 observations.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jedc.2020.104049.

References

Andolfatto, D., Gomme, P., 2003. Monetary policy regimes and beliefs. Int. Econ. Rev. (Philadelphia) 44 (1), 1–30.
Assenmacher-Wesche, K., 2006. Estimating central banks' preferences from a time-varying empirical reaction function. Eur. Econ. Rev. 50 (8), 1951–1974.
Baele, L., Bekaert, G., Cho, S., Inghelbrecht, K., Moreno, A., 2015. Macroeconomic regimes. J. Monet. Econ. 70, 51–71.
Bazzi, M., Blasques, F., Koopman, S.J., Lucas, A., 2017. Time-varying transition probabilities for markov regime switching models. J. Time Ser. Anal. 38 (3), 458–478.
Bernanke, B.S., 2007. Federal reserve communications. Speech given at the Cato Institute 25th Annual Monetary Conference in Washington, D.C. on November 14, 2007.
Bianchi, F., 2013. Regime switches, agents' beliefs, and post-World War II U.S. macroeconomic dynamics. Rev. Econ. Stud. 80 (2), 463–490.
Bikbov, R., Chernov, M., 2013. Monetary policy regimes and the term structure of interest rates. J. Econom. 174 (1), 27–43.
Brier, G.W., 1950. Verification of forecasts expressed in terms of probability. Mon. Weather Rev. 78 (1), 1–3.
Brooks, R., Harris, M.N., Spencer, C., 2012. Inflated ordered outcomes. Econ. Lett. 117 (3), 683–686.
Cochrane, J.H., Piazzesi, M., 2002. The Fed and interest rates: A high-frequency identification. Am. Econ. Rev. 92 (2), 90–95.
Cooley, T.F., LeRoy, S.F., Raymon, N., 1984. Econometric policy evaluation: note. Am. Econ. Rev. 74 (3), 467–470.
Dale, D., Sirchenko, A., 2021. Estimation of nested and zero-inflated ordered probit models. The STATA J. 21 (1) In press.
Davig, T., Leeper, E.M., 2007. Generalizing the Taylor principle. Am. Econ. Rev. 97 (3), 607–635.
Davutyan, N., Parke, W.R., 1995. The operations of the Bank of England, 1890–1908: a dynamic probit approach. J. Money Credit Bank. 27, 1099–1112.
Diebold, F., Lee, J.-H., Weinbach, G., 1994. Regime switching with time-varying transition probabilities. In: Hargreaves, C. (ed.), Nonstationary Time Series Analysis and Cointegration (Advanced Texts in Econometrics, C. W. J. Granger and G. Mizon, eds.), Oxford University Press, Oxford 283–302.
Dueker, M.J., 1999. Conditional heteroscedasticity in qualitative response models of time series: a Gibbs-sampling approach to the bank prime rate. J. Bus. Econ. Stat. 17 (4), 466–472.
Eichengreen, B., Watson, M., Grossman, R., 1985. Bank rate policy under the interwar gold standard: a dynamic probit model. Econ. J. 95, 725–745.
Epstein, E.S., 1969. A scoring system for probability forecasts of ranked categories. J. Appl. Meteorol. 8, 985–987.
Estrella, A., Hardouvelis, G.A., 1991. The term structure as a predictor of real economic activity. J. Finance 46 (2), 555–576.
Filardo, A., 1994. Business-cycle phases and their transitional dynamics. J. Bus. Econ. Stat. 12 (3), 299–308.
Frankel, J.A., Lown, C.S., 1994. An indicator of future inflation expected from the steepness of the interest rate yield curve along its entire length. Q. J. Econ. 109 (2), 517–530.
Gerlach, S., 2007. Interest rate setting by the ECB, 1999-2006: Words and deeds. Int. J. Cent. Banking 3 (3), 1–46.
Grammig, J., Kehrle, K., 2008. A new marked point process model for the federal funds rate target: methodology and forecast evaluation. J. Econ. Dyn. Control 32 (7), 2370–2396.
Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57 (2), 357–384.
Hamilton, J.D., Jorda, O., 2002. A model for the federal funds rate target. J. Political Econ. 110 (5), 1135–1167.
Harris, M.N., Zhao, X., 2007. A zero-inflated ordered probit model, with an application to modelling tobacco consumption. J. Econom. 141 (2), 1073–1099.
Hayo, B., Neuenkirch, M., 2010. Do federal reserve communications help predict federal funds target rate decisions? J. Macroecon. 32 (4), 1014–1024.
Hu, L., Phillips, P.C.B., 2004. Dynamics of the federal funds target rate: a nonstationary discrete choice approach. J. Appl. Econom. 19, 851–867.
Kauppi, H., 2012. Predicting the direction of the Fed's target rate. J. Forecast. 31, 47–67.
Kim, H., Jackson, J., Saba, R., 2009. Forecasting the FOMC's interest rate setting behavior: a further analysis. J. Forecast. 28, 145–165.
Lapp, J.S., Pearce, D.K., 2000. Does a bias in FOMC policy directives help predict intermeeting policy changes? J. Money Credit Banking 32 (3), 435–441.
Liu, J.S., Sabatti, C., 2000. Generalized Gibbs sampler and multigrid Monte Carlo for Bayesian computation. Biometrika 87 (2), 353–369.
Mishkin, F.S., 1990. The information in the longer-maturity term structure about future inflation. Q. J. Econ. 105 (3), 815–828.
Monokroussos, G., 2011. Dynamic limited dependent variable modeling and U.S. monetary policy. J. Money Credit Banking 43 (2–3), 519–534.
Müller, G., Czado, C., 2005. An autoregressive ordered probit model with application to high-frequency financial data. J. Computat. Graphical Stat. 14 (2), 320–338.
Pakko, M.R., 2005. On the information content of asymmetric FOMC policy statements: evidence from a Taylor-rule perspective. Econ. Inq. 43 (3), 558–569.
Piazzesi, M., 2005. Bond yields and the Federal Reserve. J. Political Econ. 113 (2), 311–344.
Rudebusch, G.D., 1998a. Do measures of monetary policy in a VAR make sense? Int. Econ. Rev. (Philadelphia) 39 (4), 907–931.
Rudebusch, G.D., 1998b. Do measures of monetary policy in a VAR make sense? A reply to Christopher A. Sims. Int. Econ. Rev. (Philadelphia) 39 (4), 943–948.
Russel, J. R., Engle, R., 2005. A discrete-state continuous-time model of financial transactions prices and times: the autoregressive conditional multinomial-autoregressive conditional duration model. J. Bus. Econ. Stat. 23, 166–180.
Sims, C.A., Zha, T., 2006. Were there regime switches in U.S. monetary policy? Am. Econ. Rev. 96 (1), 54–81.
Sirchenko, A., 2019. A regime-switching model for the federal funds rate target. Univ. Amsterdam Econom. Discuss. Pap. 2019/01.
Sirchenko, A., 2020. A model for ordinal responses with heterogeneous status quo outcomes. Stud. Nonlinear Dyn. Econom. 24 (1), 1–16.
Taylor, J.B., 1993. Discretion versus policy rules in practice. Carnegie-Rochester Conference Series on Public Policy 39, 195–214.
Thornton, D.L., 2005. When did the FOMC begin targeting the federal funds rate? What the verbatim transcripts tell us. Federal Reserve Bank of St. Louis Working Paper No. 2004-015B.
Vanderhart, P.G., 2000. The Federal Reserve's reaction function under Greenspan: an ordinal probit analysis. J. Macroecon. 22 (4), 631–644.
Wu, J.C., Xia, F.D., 2016. Measuring the macroeconomic impact of monetary policy at the zero lower bound. J. Money Credit Banking 48 (2–3), 253–291.