

NBER WORKING PAPER SERIES

SEMIPARAMETRIC ESTIMATES OF MONETARY POLICY EFFECTS:  
STRING THEORY REVISITED

Joshua D. Angrist  
Òscar Jordà  
Guido Kuersteiner

Working Paper 19355  
<http://www.nber.org/papers/w19355>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
August 2013

Thanks go to Jordi Gali, Ivan Werning, and seminar participants at the Bank of England, the European Central Bank, the Federal Reserve Bank of San Francisco, the NBER Summer Institute 2013, and Yale University for helpful suggestions and comments. The views expressed here are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Federal Reserve Bank of San Francisco or the Board of Governors of the Federal Reserve System, or the National Bureau of Economic Research. Early Elias provided expert research assistance.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2013 by Joshua D. Angrist, Òscar Jordà, and Guido Kuersteiner. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Semiparametric Estimates of Monetary Policy Effects: String Theory Revisited  
Joshua D. Angrist, Òscar Jordà, and Guido Kuersteiner  
NBER Working Paper No. 19355  
August 2013, Revised September 2013  
JEL No. C32,C54,E52,E58,E65

**ABSTRACT**

We develop a flexible semiparametric time series estimator that is then used to assess the causal effect of monetary policy interventions on macroeconomic aggregates. Our estimator captures the average causal response to discrete policy interventions in a macro-dynamic setting, without the need for assumptions about the process generating macroeconomic outcomes. The proposed procedure, based on propensity score weighting, easily accommodates asymmetric and nonlinear responses. Application of this estimator to the effects of monetary restraint suggest contractionary policy slows real economic activity. By contrast, the Federal Reserve's ability to stimulate real economic activity through monetary expansion appears to be much more limited. Estimates for recent financial crisis years are similar to those for the earlier, pre-crisis period.

Joshua D. Angrist  
Department of Economics, E17-226  
MIT  
77 Massachusetts Avenue  
Cambridge, MA 02139  
and NBER  
angrist@mit.edu

Guido Kuersteiner  
Department of Economics  
Tydings Hall  
The University of Maryland  
College Park, Maryland 20742  
gk232@Georgetown.edu

Òscar Jordà  
Economic Research, MS 1130  
Federal Reserve Bank of San Francisco  
101 Market St.  
San Francisco, CA 94105  
and University of California, Davis  
oscar.jorda@sf.frb.org

“Our credibility will ultimately be judged by how we do on both of these mandates, not just the price mandate,” Mr. (Charles) Evans said Tuesday night. “I think we will be judged very badly” if officials do not act forcefully to reduce unemployment and instead, he said, “worry obsessively” about inflation.

“There’s little more that we can do,” Mr. (Jeffrey) Lacker said of monetary policy. “There’s little more that we can contribute to growth.”

- *The New York Times*, April 2, 2013

Congressman Goldsborough: You mean you cannot push a string.

Governor Eccles: That is a good way to put it, one cannot push on a string. We are in the depths of a depression and ... beyond creating an easy money situation through reduction of discount rates and through the creation of excess reserves, there is very little if anything that the reserve organization [Federal Reserve Board] can do toward bringing about recovery. I believe that in a condition of great business activity that is developing to a point of credit inflation, monetary action can very effectively curb undue expansion.

- Testimony before the House Committee on Banking and Currency, March 18, 1935.

## 1 Introduction

The string metaphor is an enduring feature of the debate over monetary policy: increasing borrowing costs may reliably slow an expansion, but cheap capital need not stimulate investment in a downturn. What does the historical record say regarding the possibly different macroeconomic effects of monetary contraction and expansion? As many have recognized, this question is easy to ask but hard to answer. The identification challenge in this context arises from the fact that policy changes are rarely isolated from other economically important developments, including, perhaps, anticipated changes in economic conditions. If these changes are related to the outcome variables of interest, one subset of time series observations likely provides a poor control for another.

Many contemporary investigations of macro policy rely on structural models of economic behavior to solve this fundamental identification problem (see, for example, the survey by Christiano, Trabandt and Walentin, 2011). This approach, typically cast in a dynamic structural general equilibrium (DSGE) framework, begins with a model of the macroeconomy that is meant to mimic the time series behavior of key macro variables. In addition to theoretical predictions, DSGE models generate a system of linear (or linearized) difference equations that provide the basis for empirical work. These equations can be interpreted as vector autoregressions (VARs) with an associated set of coefficient or error-covariance restrictions (as in Bagliano and Favero, 1998; Christiano, Eichenbaum and Evans, 1999; and many others). The heart of the DSGE approach uses a model of the entire economy to isolate shocks that identify causal policy effects. The validity of the resulting causal inferences therefore turns in part on how accurately economic models describe the macroeconomy.

An alternative strategy, inspired by the landmark Friedman and Schwartz (1963) volume, tries to identify policy shocks through a close reading of the Federal Reserve’s Open Market Committee (FOMC) meeting minutes. Romer and Romer (1989) is the first in a series of influential contemporary studies in this mold. A drawback of the narrative approach is the subjective manner in which shocks are identified. Moreover, some of the putatively random policy shifts identified in Romer and Romer (1989) may be correlated with economic variables (an argument fleshed out in Hoover and Pérez, 1994; see Romer and Romer, 1994 for a rebuttal).

This paper outlines a new route to causal inference for monetary policy effects, sidestepping some of the difficulties encountered in structural and narrative-based efforts. The defining feature of our approach is a laser-like focus on the policy-making process. In contrast with the narrative approach, which also focuses on Fed decision-making, our analysis of Fed behavior is more formal and data-driven. Our solution to the policy evaluation problem starts with the presumption that, conditional on market-derived statistics that embed optimal forecasts of future outcomes and anticipated policy moves, along with a small set of institutional and economic variables, policy variation can be used to identify causal effects. This assumption allows us to quantify the causal effect of policy changes in an environment of stable expectations and goals. At the same time, our empirical strategy easily accommodates nonlinear effects, while distinguishing the effects of monetary easing from those of tightening.

The selection-on-observables framework outlined here is founded on strong identifying assumptions, but also provides a natural starting point for time series causal inference. In the absence of purposefully designed experiments or naturally occurring quasi-experimental shifts, it’s hard to see how one can do better than to use the policy variation at hand. Our focused approach limits the task of model specification and robustness checking to the formulation and testing of a model of the policy determination process. The selection-on-observables assumption also buys us strong testable restrictions that can be used to assess the plausibility of causal claims. The principal econometric question that arises in our context is how to exploit selection-on-observables identification in a manner that imposes minimal auxiliary assumptions and facilitates specification testing.

Our econometric policy model describes the probability of federal funds rate target changes conditional on market statistics, past policy choices, lagged outcomes, and a few other controls. The resulting set of conditional distributions defines a function we call the *policy propensity score*. Monetary policy rules have long been studied in macroeconomics; see, for example, Clarida, Galí and Gertler (2000), Woodford (2001) and Galí and Gertler (2007) for recent contributions. Propensity score methods, introduced by Rosenbaum and Rubin (1983), have proven useful for cross-sectional causal inference (see, e.g., Dehejia and Wahba (1999) and Heckman, Ichimura and Todd(1998)). In a pair of papers closely related to this one, Angrist and Kuersteiner (2004, 2011), adapt the propensity score framework to the problem of time series causality testing of the sort discussed by Granger (1969) and Sims (1972). We extend this framework here, deriving flexible, easy-to-compute propensity score estimators of the causal effects of a dynamic multinomial treatment. These semiparametric estimators are then used to assess the impact of

monetary policy before and since the Great Recession. The main payoff to our approach is the ability to go directly from the policy process to causal effects on outcomes. The resulting causal estimates are valid for all processes generating outcomes, nonlinear and complex as they might be, while allowing distinct assessments of the impact of tight and easy money.

The first task on our empirical agenda is the construction of a credible model for the policy propensity score. To that end, we follow Kuttner (2001), Faust, Swanson and Wright (2004), and Piazzesi and Swanson (2008) in using market-based measures of anticipated policy. Specifically, we rely on Piazzesi’s (2005) model linking Federal Reserve policy actions with asset prices. This link is used to justify a model for target rate changes as a function of the price of federal funds rate futures contracts. Market-based predictions of policy actions provide a low-dimension aggregator of publicly available information. This in turn leads to parsimonious policy models well suited to a data-poor time series setting.

Our investigation of monetary policy effects replicates findings from earlier work while uncovering some that are new. Echoing the findings of Christiano, Eichenbaum and Evans (1996, 1999), among others, our results suggest contractionary monetary policy slows real economic activity, reducing employment as well as inflation. At the same time, in contrast with a number of earlier studies (such as those reviewed in Christiano, Eichenbaum and Evans, 1999), the semiparametric estimation strategy developed here suggests Fed efforts to support the real economy have been disappointing. The ease with which this asymmetric response is revealed is an attractive feature of our approach. Finally, motivated by interest in the Fed’s attempts to stimulate the economy during the Great Recession, we compare responses calculated using a sample that ends in mid-2005 with results from a sample running through 2008, when the federal funds target rate hit zero. We also compute estimates for the latter period only. Our conclusions regarding the Fed’s limited ability to boost real economic activity stand under both variations.

## 2 Potential Outcomes and Macro Causal Effects

### 2.1 Conceptual Framework

The economy is described by the vector,  $\chi_t = (x_t', y_t', D_t')'$ , where  $y_t$  is a  $k_y$ -vector of outcome variables,  $D_t$  is a vector of policy variables that takes on values  $d_0, \dots, d_J$ , and  $x_t$  is a vector of contemporaneous covariates.

Policy is determined by lagged economic conditions, lagged policy choices, and covariates, combined in the vector  $z_t = f[x_t, \chi_{t-1}]$ . The policy regime is indexed by a parameter,  $\psi$ , which takes values in a parameter space  $\Psi$ . In addition, policymakers are assumed to react to idiosyncratic information or taste variables, represented by the scalar  $\varepsilon_t$ , that we don’t get to see. The realized policy  $D_t$  is determined by both observed and unobserved variables according to  $D_t = D(z_t, \psi, \varepsilon_t)$ . For identification purposes, we assume that  $\varepsilon_t$  is independent of potential outcomes. This is reminiscent of the recursive ordering proposed by Christiano, Eichenbaum and Evans (1996, 1999), but our approach requires no description

of the structural process connecting  $y_t$  with policy choices or expected future values of  $\chi_t$ .

Our framework for causal inference builds on the notion of potential outcomes. Potential outcomes describe realizations of  $y_t$  that arise in response to a hypothetical change in monetary policy. The potential outcomes concept originated in experimental studies where the investigator has control over the assignment of treatments, but is now widely used in observational studies. Although potential outcomes most commonly appear in studies looking at the causal effects of a binary treatment or policy intervention, the idea is easily extended to ordered discrete or continuous interventions (see, e.g., Angrist and Imbens, 1995).

The definition of potential outcomes used here comes from Angrist and Kuersteiner (2011):

**Definition 1** For fixed  $t, l$ , and  $\psi$ , potential outcomes  $\{y_{t,l}^\psi(d); d \in \mathcal{D}\}$  are defined as the set of values the observed outcome variable  $y_{t+l}$  would take on if  $D_t = D(z_t, \psi, \varepsilon_t) = d$ , for all possible policy choices  $d$ ,  $d \in \{d_0, \dots, d_j, \dots, d_J\}$ .

The vector of potential outcomes includes the observed outcome,  $y_{t+l} = y_{t,l}^\psi(D_t)$ , as well as counterfactual outcomes describing the consequences of policy choices not taken. The causal effect of a policy change is defined as the difference  $y_{t,l}^\psi(d_j) - y_{t,l}^\psi(d_0)$ , where  $d_j$  indicates an intervention and  $d_0$  indicates a benchmark policy. This notation makes a conceptual distinction between changes to the policy regime (indicated by changes in  $\psi$ ) and policy changes within the same regime, though only the latter are identified in our framework. Although the notation introduced here is nonstandard in a macro setting, the notion of macroeconomic causal effects determined by counterfactual states has a long history. Cochrane (1994), for example, quotes Kareken and Solow (1963): "... One cannot deduce conclusions about the effects of monetary policy or about their timing without making some hypothesis, explicit or implicit, about what the course of events *would have been* had the monetary authorities acted differently." (emphasis added)

In a macroeconomic context, potential outcomes can be generated by a DSGE structure. Suppose, as in Cochrane (1994), that output is determined by

$$y_t = a_{y\varepsilon}\varepsilon_t + a_{yD}D_t + a_{y\eta}\eta_t, \tag{1}$$

where the  $a$ . are parameters and  $\eta_t$  is an i.i.d. shock (The mapping between  $a$  and  $\psi$  may be obtained by linearizing a non-linear DSGE model.). Cochrane's monetary policy rule is given by

$$D_t = a_{Dy}y_{t-1} + \varepsilon_t \tag{2}$$

The resulting moving average representation for  $y_t$  can be written

$$y_t = \sum_{j=0}^{\infty} c_{y\varepsilon,j}\varepsilon_{t-j} + \sum_{j=0}^{\infty} c_{y\eta,j}\eta_{t-j}.$$

Here,  $c_{y\varepsilon,j}$  is the impulse response function for output in response to policy shocks. Suppose, that instead of the realized policy,  $D_t$  is set to  $d_j$ . The potential outcome induced by this manipulation is:

$$y_{t,l}(d_j) = c_{y\varepsilon,j}(d_j - a_{Dy}y_{t-1}) + \sum_{j=0, j \neq l}^{\infty} c_{y\varepsilon,j} \varepsilon_{t+l-j} + c_{y\eta}(L) \eta_t. \quad (3)$$

This tell us what we can expect to see in response to such a change, in a world where the path of shocks is otherwise unchanged. The associated causal effect of this policy change is  $y_{t,l}^\psi(d_j) - y_{t,l}^\psi(d_0) = c_{y\varepsilon,j}(d_j - d_0)$ . Thus, causal effects in our framework match those determined by a VAR in a linear, constant-effects world.

Individual causal effects can never be observed since the real world gives us on only one realization. We therefore focus on average causal effects. Let  $Y_{t,L} = (y'_{t+1}, \dots, y'_{t+L})'$  and define the vector of potential outcomes up to horizon  $L$  by  $Y_{t,L}^\psi(d) = (y_{t,1}^\psi(d), \dots, y_{t,L}^\psi(d))$ . Potential outcomes determine observed outcomes as follows:

$$Y_{t,L} = \sum_{d \in \mathcal{D}} Y_{t,L}^\psi(d) 1\{D_t = d\}. \quad (4)$$

Average policy effects are given by

$$E \left[ Y_{t,L}^\psi(d_j) - Y_{t,L}^\psi(d_0) \right] \equiv \theta_j, \quad (5)$$

where  $\theta_j = (\theta'_{1,j}, \dots, \theta'_{L,j})'$  describes the response of  $y_t$  to policy  $d_j$  at horizons 1 to  $L$ . The effects of all possible policy changes are summarized by  $\theta = (\theta'_1, \dots, \theta'_J)'$ , a vector of dimension  $k = k_y \bullet L \bullet J$ , with  $k_y$  the number of outcome variables,  $L$  the horizon of interest, and  $J + 1$  the number of policy options.  $\theta$  describes an average generalized impulse response function for all possible policy choices. In contrast with traditional impulse response analyses in empirical macro, however, the impulse response function estimated here can be both asymmetric and nonlinear.

Potential outcomes for counterfactual policy choices are unobserved, so the expectation in (5) cannot be estimated directly. The variation that identifies causal relationships in our framework is characterized by a conditional independence assumption, also known as *selection on observables*:

**Condition 1** *Selection on observables:*

$$y_{t,l}^\psi(d_j) \perp D_t | z_t \text{ for all } l \geq 0 \text{ and for all } d_j, \text{ with } \psi \text{ fixed; } \psi \in \Psi.$$

In Cochrane's example, the policy function, indexed by  $\psi$ , affects the impulse response parameters,  $c_{y\varepsilon}$ , through the parameter  $a_{Dy}$ . Cochrane (1994) shows that  $c_{y\varepsilon}$  may change as fed behavior changes. Our conditional independence assumption focuses on variation in policy interventions while holding the policy regime fixed, after conditioning on observables,  $z_t$ . Leeper and Zha's (2003) notion of *modest policy interventions* captures the same idea. In the simple model described by (1) and (2), selection on

observables requires serially independent  $\varepsilon_t$ , independent of  $\eta_t$ .

Using Equation (4) and Condition 1, we can write the average policy effect conditional on  $z_t$  in terms of observable distributions as:

$$E \left[ Y_{t,L}^\psi(d_j) - Y_{t,L}^\psi(d_0) | z_t \right] = E [Y_{t,L} | D_t = d_j, z_t] - E [Y_{t,L} | D_t = d_0, z_t]. \quad (6)$$

Although cast in terms of in-principle observable conditional means, in applications with a high-dimensional conditioning set involving continuous random variables, estimation of these conditional expectations is empirically demanding. The estimation problem is simplified by use of a parametric model for the policy function. Angrist and Kuersteiner (2011) call this model the *policy propensity score*.

The policy propensity score is  $P(D_t = d_j | z_t) = p^j(z_t, \psi)$ , where  $p^j(z_t, \psi)$  is a flexible parametric model with parameters determined by the policy regime. Average causal effects can then be estimated using the fact that Condition 1 implies

$$E [Y_{t,L} 1 \{D_t = d_j\} | z_t] = E \left[ Y_{t,L}^\psi(d_j) | z_t \right] p^j(z_t, \psi). \quad (7)$$

Solving (7) for  $E \left[ Y_{t,L}^\psi(d_j) | z_t \right]$  and integrating over  $z_t$  allows us to write

$$\theta_j = E \left[ Y_{t,L}^\psi(d_j) - Y_{t,L}^\psi(d_0) \right] = E \left[ Y_{t,L} \left( \frac{1 \{D_t = d_j\}}{p^j(z_t, \psi)} - \frac{1 \{D_t = d_0\}}{p^0(z_t, \psi)} \right) \right]. \quad (8)$$

This weighting scheme was first used to estimate population means in non-random samples by Horvitz and Thompson (1952) and adapted for the estimation of causal effects of cross-sectional Bernoulli treatments by Hirano, Imbens, and Ridder (2003). In cross-sectional studies of causal effects of Bernoulli interventions, (8) is known as an *average treatment effect*. Our setup allows for multinomial or ordered treatments.<sup>1</sup>

The estimand described by (8) is similar to that approximated using local linear projections in Jordà (2005), though here no approximation is required. The estimand can also be related to the nonlinear impulse response function introduced by Gallant, Rossi and Tauchen (1993). The latter is based on estimation of  $E[y_{t+l} | x_t]$  where  $x_t = (y_t, \dots, y_{t-p})$  and  $y_t$  is assumed to be a Markov process. An impulse response function can then be defined as a marginalized version of  $E[y_{t+l} | x_t^+] - E[y_{t+l} | x_t]$  where  $x_t^+$  perturbs  $y_t$  by a constant. Although  $E[y_{t+l} | x_t]$  is in principal nonparametrically identified, extrapolation to counterfactual  $E[y_{t+l} | x_t^+]$  in the Gallant, Rossi and Tauchen (1993) framework is likely to require a model for the conditional expectation of outcomes.

Our approach leans on parametric policy models but requires no functional form assumptions for the

---

<sup>1</sup>With Bernoulli treatments, the formulation in (8) reflects Rosenbaum and Rubin's (1983) propensity score theorem, which says that if potential outcomes are independent of treatment conditional on covariates, they're also independent of treatment conditional on the propensity score. Imbens (2000) notes briefly that Horvitz-Thompson weighting applies to multinomial treatments as well.

outcome process,  $y_t$ . We use institutional knowledge and economic reasoning to guide the choice of  $z_t$ , and specification testing to assess our model for  $p^j(z_t, \psi)$ . We need not define or estimate structural innovations for the policy process, nor develop, solve, or simulate a model for the joint process governing outcomes  $y_t$ , policy variables  $D_t$ , and covariates  $z_t$ . The estimator based on (8) is an easily constructed weighted average, for which inference is straightforward.

## 2.2 Estimation

Inverse probability weighting estimators can be written as simple weighted averages of the vector of future outcomes  $Y_{t,L} = (y'_{t+1}, \dots, y'_{t+L})'$ , with weights formed from:

$$\delta_{t,j}(\psi) = \frac{1\{D_t = d_j\}}{p^j(z_t, \psi)} - \frac{1\{D_t = d_0\}}{p^0(z_t, \psi)}.$$

In a correctly specified model, these weights have mean zero and are uncorrelated with  $z_t$ . To ensure this is true in finite samples and under possible misspecification, we weight with the residuals from a regression of  $\delta_{t,j}(\hat{\psi})$  on  $z_t$  and a constant. This produces the weighting function,  $\hat{h}_t = Y_{t,L}(\delta_{t,j}(\hat{\psi}) - \hat{\delta}_{t,j})$ , where  $\hat{\delta}_{t,j}$  is the relevant fitted value. The average causal response vector is then estimated as the sample average

$$\hat{\theta} = T^{-1} \sum_{t=1}^T \hat{h}_t. \quad (9)$$

The estimator  $\hat{\theta}$  solves

$$\hat{\theta} = \arg \min_{\theta} \left( T^{-1} \sum_{t=1}^T \hat{h}_t - \theta \right)' \Omega^{-1} \left( T^{-1} \sum_{t=1}^T \hat{h}_t - \theta \right), \quad (10)$$

a minimum distance objective function. Given this problem has an equal number of parameters and moment conditions, the choice of  $\Omega$  leaves the solution to (10) unaffected. Sometimes, however, it's interesting to restrict  $\theta$ . For example, to facilitate comparisons with VAR's, symmetric responses can be estimated by assuming that the effect of  $d_j$  and  $-d_j$  are of the same magnitude but opposite in sign. More generally, we estimate under restrictions  $\theta = \theta(\alpha)$ , where  $\alpha$  is a reduced set of free parameters, by solving:

$$\hat{\alpha} = \arg \min_{\alpha} \left( T^{-1} \sum_{t=1}^T \hat{h}_t - \theta(\alpha) \right)' \Omega^{-1} \left( T^{-1} \sum_{t=1}^T \hat{h}_t - \theta(\alpha) \right). \quad (11)$$

The optimal  $\Omega$  for this over-identified scenario is the spectral density matrix of  $\hat{h}_t$  at zero frequency, which can be estimated as detailed in Newey and West (1994). Asymptotic approximations to the sampling distribution of  $\hat{\theta}$  or  $\hat{\alpha}$  account for the fact that  $\psi$  is estimated in a first stage. The relevant limiting distributions for estimators and test statistics are derived in the appendix.

### 3 A Propensity Score for Monetary Policy Interventions

The federal funds market is an interbank loan market intended for the management of reserve requirements; the rate for overnight loans in this market, known as the federal funds rate, provides a benchmark for securities across the risk and maturity spectrum. Monetary policy targeted the level of the federal funds rate until mid-December, 2008, when the fed funds rate was set to trade between 0 and 0.25%. With no room to lower rates further, the Fed turned to other tools, such as large scale asset purchases. We focus here on the pre-2009 policy era, going back to July 1989. Because FOMC meetings are very nearly a monthly occurrence, we work with monthly data.

#### 3.1 The Federal Funds Rate Targeting Regime

Since February 1994, fed funds targets have been announced after each FOMC meeting, eight times a year.<sup>2</sup> Changes in the target usually come in 25 basis-point increments in a [-0.50%, 0.50%] interval, though the target was twice changed by 75 basis points in our sample period. Of the 78 target rate changes in our sample, 23 were outside of an FOMC meeting. Most of these happened before February 1994. On three occasions there was more than one change in a given month, in which case we sum them.<sup>3</sup> The space of possible policy choices is defined here to be  $\{-0.50\%, -0.25\%, 0\%, 0.25\%, 0.50\%\}$ , where the  $\pm 0.50\%$  events include the larger changes.

Our sample is determined in part by the availability of policy predictors derived from financial derivatives based on the federal funds rate. One such derivative, federal funds rate futures (FFF) started trading in earnest in 1989 (the market was established earlier, in 1988). We consider two sample end-points. July 2005 marks the last in a series of increases in the fed funds target. From then on, the target rate remained at 5.25% until September 2007 before the Great Recession began. We use July 2005 as a first sample end-point for estimation of the propensity score and evaluate policy responses up to 24 months ahead on data running to July 2007. The target fell gradually thereafter, until December 2008 when it hovered between zero and 0.25%, marking the end of conventional monetary policy. The second sample end-point for propensity score estimation therefore extends through the end of 2008 with an additional 24 months ending in December 2010 used to estimate policy responses. Finally, we also experiment separately with data from the Great Recession period only. This is a small sample and it imposes some limits to the experiments we can conduct that we will discuss below.

#### 3.2 Policy Predictors for the Propensity Score

Efficient markets price futures contracts using all available information. This motivates students of monetary policy to define policy shocks as deviations from the optimal predictions implicit in asset prices.

---

<sup>2</sup>The FOMC comprises the Fed Chairman and six other Board Governors, the president of the Federal Reserve Bank of New York and a rotating pool of four presidents from the remaining eleven regional Federal Reserve Banks.

<sup>3</sup>The relevant months are December 1990, December 1991, and January 2001.

Cochrane and Piazzesi (2002), for example, use changes in the Eurodollar rate around meeting dates to define target rate surprises, while Thapar (2008) uses T-Bill futures in a similar manner. Kuttner (2001), Faust, Swanson and Wright (2004), Gürkaynak, Sack and Swanson (2005, 2007), Bernanke and Kuttner (2005) and Wingender (2011) construct monetary policy shocks from financial derivatives that price the federal funds rate directly.<sup>4</sup>

Federal funds rate derivatives include a futures contract on the effective federal funds rate and an options contract on these futures (though the latter started only in 2003). Futures contracts refer to calendar-month averages of the effective federal funds rate published by the New York Fed, with spot, and one- through five-month contracts. We use these derivatives to predict target changes, implicitly defining policy surprises as deviations from market-based forecasts of Federal Reserve behavior.

The intuitive notion that futures prices provide an optimal policy forecast can be made rigorous using Piazzesi's (2005) term structure model. Denote the information available to policy makers at time  $t$  by  $\zeta_t$ . Piazzesi shows that bond yields and related derivatives likewise depend on  $\zeta_t$ . Under the additional assumption that pricing functions are invertible, we can replace  $\zeta_t$ , which may be only partially observable, with a vector of observed asset prices  $z_t$ . This theoretical argument is fleshed out in the appendix.

As a practical matter, our analysis distinguishes between months with scheduled FOMC meetings ("meeting months") and months without. In meeting months, we construct  $s_t^1$ , the difference between the price of the futures contract expiring during the same month as the meeting and the current target rate, both observed at market close on the day before the meeting. In non-meeting months, we construct  $s_t^0$ , the difference between the price of a one-month-ahead contract in that month and the target rate in effect on the last day of the previous month.

With a few exceptions, our predictor for meeting months is  $s_t^1$ , while in non-meeting months,  $s_t^0$  is used. In constructing  $s_t^1$  for months in the pre-1994 era, before target rate changes were announced, our coded announcement date is delayed by one day relative to the later period. Before 1994, the market became aware of a target change only through a reading of the open market operations implemented by the New York Fed's Trading Desk. These operations take place at the beginning of the trading day and hence are observed the day after a meeting, which generally concludes after the close of the market. This nuance affects the construction of  $s_t^1$  only.

In a few instances, target rate changes in meeting months preceded meetings, with an additional change or no change at the meeting, as in February, March, and August 1991. In such situations, the predictor is taken to be  $s_t^0$  rather than  $s_t^1$  and enters as its own covariate. The active futures variable for any given month is denoted  $FFF_t$ , equal to either  $s_t^0$  or  $s_t^1$ , as described above.

Because target changes are naturally ordered in 0.25% increments over the range  $\pm 0.50\%$ , we model the policy propensity score with an ordered probit specification. Hamilton and Jordà (2002) and Scotti (2011) likewise use ordered probit to model federal funds rate target changes. The dependent variable

---

<sup>4</sup>Krueger and Kuttner (1996), Kuttner (2001), Piazzesi and Swanson (2008), Gürkaynak et al. (2005, 2007) and Hamilton (2009) discuss the efficiency of the market for federal funds futures.

is the change in the target rate during month  $m + 1$ , irrespective of whether an FOMC meeting was announced or whether the target was changed. For the few occasions where more than one target change occurred in the same month, recall that we use the accumulated monthly change.

We allow the coefficients of the propensity score to vary after August 2006 to capture a window that includes the financial crisis with about a one year lead time.<sup>5</sup> Since there were no instances of changes prior to meetings after 2001, the predictor for that scenario only enters in the pre-crisis data.

In addition to controlling for market expectations through fed funds futures prices, we include the unemployment rate, such as might appear in a conventional monetary policy rule. The unemployment rate is available at a monthly frequency and is a natural substitute for output gap measures commonly used with quarterly data. Moreover, our "Taylor Rule" specification can be motivated by results of Blanchard and Galí (2010) and Galí (2011) who show that the optimal monetary policy rule in a New Keynesian model with real-wage rigidities depends both on inflation and unemployment. Finally, the specification of the propensity score includes a number of terms designed to capture a variety of calendar effects. These include a dummy variable  $FOMC_t$  indicating months with a scheduled FOMC meeting, the target change in the previous month, the target rate change in the previous month interacted with FOMC, a scale factor that accounts for when within the month the FOMC meeting is scheduled, and a set of monthly seasonal dummies. We also include the variable  $CRISIS_t$ , a dummy that takes the value of 1 starting August 2006 to capture a window that includes the financial crisis with about a one year lead time. Finally, the model includes dummies for the Y2K event and the September 11, 2001 attacks.<sup>6</sup>

Table 1 reports average marginal effects of the effects of predictors on the likelihood of a 0.25% increase in the target. These estimates use the sample through July 2005 for the pre-crisis sample (July 1989 to July 2005) and through December 2008 for the full sample (July 1989 to December 2008). Since responses are calculated by shifting the outcome variable up to two years forward, we note that the end dates for the outcome samples are July 2007 and December 2010, respectively.

Estimates of a benchmark Taylor-type specification that predicts target rate changes with inflation and unemployment alone are reported in columns (1) and (2). These are labeled  $OP_{T1}$  and  $OP_{T2}$  and use the pre-crisis sample. Columns (5) and (6), report estimates for  $OP_{T1}$  and  $OP_{T2}$  using the full sample. Broadly speaking, the estimates show that both variables affect policy largely as expected, though the negative unemployment effect is stronger than the very small positive inflation effect. The latter is not statistically significant. The Taylor model estimates shown in columns (1) and (5) (labeled  $OP_{T1}$ ) use same-month measures only, while the estimates reported in columns (2) and (6) (labeled  $OP_{T2}$ ) are from models that add inflation and unemployment lags, the size of the last target change, and seasonal and scheduling dummies.<sup>7</sup>

---

<sup>5</sup>The last target increase is dated June 2006. August 2007 marks the first in a sequence of target reductions that extend to the end of the sample.

<sup>6</sup>The scale factor is defined as  $\kappa/(\kappa - t)$  where  $\kappa$  is the number of days in a given month and  $t$  is the day of the month when the FOMC meeting is scheduled.

<sup>7</sup>Detailed variable definitions and sources appear in Appendix B.

Columns (3) and (4) for the short sample and columns (7) and (8) for the long sample, labeled  $OP_{F1}$  and  $OP_{F2}$  report estimates from specifications that include  $FFF_t$  terms that differ in meeting and non-meeting months, and in the pre-crisis and full samples. The results indicate that market-based factors are better predictors of target rate changes than the combination of inflation and unemployment and their lags in a Taylor specification. The pseudo- $R^2$ s in columns (3) and (7) ( $OP_{F1}$ ), based on estimates of the policy propensity score using these factors but excluding inflation and unemployment terms are virtually identical to the pseudo- $R^2$ s from the specifications augmented with these Taylor model terms and reported in columns (4) and (8) ( $OP_{F2}$ ). Neither inflation nor unemployment marginal effects are significantly different from zero when estimated in the more elaborate models.

Fitted values from the full policy score model (8) seem to track realized shifts well over the course of the business cycle. This can be seen in Figure 1, which plots actual and predicted target changes (i.e., the expected target change conditional on regressors in the policy propensity score). Predictions were computed using the  $OP_{F2C}$  estimates from column (8) of Table 1. The figure also shows the time series of Industrial Production (IP) growth to mark cyclical fluctuations.

An important diagnostic for our purposes looks at whether lagged macro aggregates are independent of policy changes conditional on the policy propensity score. In other words we would like to show that the policy shocks implicitly defined by our score model look to be "as good as randomly assigned". Angrist and Kuersteiner (2011) develop semiparametric tests that can be used for this purpose. Panel A of Table 2 reports test results for the null hypothesis of orthogonality between the policy innovation,  $\{D_t = d_j\} - p^j(z_t, \psi)$ , and conditioning variables in the model. Panel B looks at the correlation between innovations and lagged outcomes. Both panels show p-values for joint (3 *df*) tests of the null hypothesis that the score model indicated in the column headings is an adequate specification of the conditional probability that the target rate is unchanged, rises by a quarter point, or falls by a quarter point (these are the shocks analyzed below). The table also shows p-values for joint orthogonality of an aggregate market based factor  $FFF_t$ . We consider linear and quadratic terms of  $FFF_t$  ( $2 \times 3 = 6$  *df*).

Test results for the simple Taylor model (reported in columns labelled  $OP_{T1}$ ) show substantial correlation with economic variables, including the lagged federal funds rate. The addition of controls for inflation and unemployment lags, the size of the last target change, and seasonal and scheduling dummies generates better results, although these p-values (reported in columns labelled  $OP_{T2}$ ) still show some evidence of correlation between  $FFF_t$  and the average federal funds rate with policy innovations, especially in the longer sample. Test results for the model with  $FFF_t$  alone (reported in columns labelled  $OP_{F1}$ ) all pass, contrary to those for the expanded Taylor model,  $OP_{T2}$ . Similarly, test results for the model that includes both  $FFF_t$  and inflation and unemployment terms (reported in columns labeled  $OP_{F2}$ ) offer no evidence against the hypothesis of random policy innovations. In what follows, we proceed using the p-score model based on  $OP_{F2}$ .

## 4 Dynamic Policy Effects

The yield curve is the proximate channel through which target rate changes affect inflation and the real economy. We therefore begin with an analysis of policy effects on the yield curve, specifically, the federal funds, 3-month T-Bill, and 2- and 10-year T-Bond rates. In addition to the yield curve, we look at effects on inflation measured by the change in 100 times the log of the Consumer Price Index (CPI), the response in the change of 100 times the log of the Industrial Production (IP) index, and the response in the change of the unemployment rate. Policy responses refer to the percentage point change in the outcome variable measured from the month of the policy intervention out to the relevant horizon.

Our analysis shows the impact of 25 basis point changes, up and down, out to a horizon of 24 months. Figures 2, 3 and 4 plot these estimated responses, constructed using propensity score model  $OP_{F2/C}$  in columns (4) and (8) of Table 1. The figures also show 90% confidence bands.

The federal funds rate responds more sharply to increases in the target rate than to decreases, as can be seen in Figure 2. A 25 basis point (bps) increase in the target appears to spark a sequence of further changes that induces a peak increase in the federal funds rate close to 1 percentage point (pp) after about a year, then falling to half a percentage point after two years. This pattern is similar to that found in VAR-type estimates (e.g., Figure 3 in Christiano et al. 1999, when cumulated). By contrast, a 25 bps reduction lowers the federal funds rate by -0.50 pp one year out, a decline that's sustained after two years.

Estimated causal effects of a change in target yields are shown in Figure 3. As we might expect, rate increases move through the yield curve with diminished intensity as maturities lengthen. A similar pattern appears in estimates reported by Cochrane and Piazzesi (2002). The estimated yield curve response to a rate decrease goes the other way and is somewhat more muted, as can be seen in the right-hand column of the figure. This relatively flat response is consistent with the flatter response of the federal funds rate in the bottom of Figure 2. The relative insensitivity of the yield curve to target rate decreases provides a possible explanation for weak transmission of accommodative monetary policy changes to macro aggregates.

Responses of macroeconomic aggregates to monetary policy changes appear in Figure 4. The top row displays the responses of the change in 100 times the log of the price level; the second row reports effects on 100 times the change in the log of IP; the third row shows effects on the change in the unemployment rate. Target rate increases begin to reduce inflation after about 18 months. After two years, a quarter point increase in the target is estimated to have reduced prices by about a quarter point, equivalent to a reduction of just over a tenth of a point in the annual inflation rate. Interestingly, these results show no evidence of a "price puzzle," that is, a short-run increase in inflation when rates rise. The price puzzle is a common finding in VAR-based estimates of the effects of monetary shocks (see, e.g., Sims, 1992)

Industrial Production is largely invariant to a rate increase in the first post-change year, but then falls to a net decline of about 1.5 pps after two years. This is roughly a 0.75 pp decrease in annual growth

rates. The unemployment rate response essentially mirrors this, with a total increase of about a quarter point after two years. Estimated responses of IP and the unemployment rate to target rate reductions emerge only slowly and are modest when they do. This difference, documented in the right-hand column of Figure 3, seems likely to reflect the relatively muted response of the federal funds rate and the yield curve to reductions in the target rate. Although the effect of target rate changes is about equal, up or down, when looking two years out, the fed funds rate and yield curve are markedly more sensitive to rate increases than to rate decreases at intermediate horizons. Differences in the responses to up and down changes therefore seem likely to reflect a weaker "first stage" for the policy intervention, though we can't rule out differences in policy sensitivity on the upside and downside as well. This strong asymmetry in yield curve and macro aggregate responses to US monetary policy shifts echoes findings in Hamilton and Jordà (2002) and Angrist and Kuersteiner (2011), but does not feature in most VAR-based estimates.

### **Other Comparisons and Context**

In an influential study of the effects of monetary policy shocks on the yield curve and macro variables, Cochrane and Piazzesi (2002) report estimates of policy effects on the yield curve similar to ours. On the other hand, their results show little effect of policy changes on prices, while suggesting employment increases after a rate increase. The yield curve effects reported here are stronger than the VAR-based responses reported in Christiano, Eichenbaum and Evans (1996, 1999).

Faust, Swanson and Wright (2004) use policy-induced changes in federal funds futures price to quantify policy shocks. Their VAR-based estimates of the effect of a positive 25 basis point surprise show price decreases similar to those reported here. The corresponding estimated effects on output line up less well, however, with a mixture of positive and negative effects.

As a theoretical matter, monetary models with nominal rigidities, information asymmetries, menu costs, or lending constraints typically imply asymmetric responses to monetary policy interventions. For example, Cover (1992) and DeLong and Summers (1988) argue that contractionary monetary policy affects real variables more than expansionary policy does. Using international data, Karras (1996) finds strong evidence of asymmetry in the effects of monetary policy on output using European data. These papers are consistent with Keynes' (1936) observations on the role of sticky wages in business cycles (see Ravn and Sola, 2004 for a review of the relevant history of thought in this context). More recently Tenreyro and Thwaites (2013), using Romer and Romer (2004) monetary policy shocks, report that monetary policy effectiveness declines in recessions.

## **5 Crisis Intervention**

Did the Great Recession change the effects of monetary policy? This section addresses this question by first extending the sample to cover target rate changes through the end of 2008, with outcomes measured through the end of 2010 and then by analyzing policy changes for the early crisis period only (October

2006 to December 2008), with outcomes measured through December 2009 (one year earlier than the full-sample analysis as we limit the horizon to 12 months for the crisis sample).

Figure 5 shows the estimated response of the federal funds rate in the longer sample (the first of our two experiments). These estimates are similar to those shown in Figure 2, with the response to a target increase peaking slightly above 1 pp after 18 months, falling to a little over half a point two years out. The estimated response to a target decrease is slightly less persistent than in the shorter sample, ending the two-year horizon with a decline of only a quarter point rather than a half.

Estimated term rate responses are plotted in Figure 6. These show a slightly sharper yield curve response to rate increases, while the response to target rate reductions is, if anything, even more muted than that for the pre-crisis sample. The estimated effects in panels on the right hand side of Figure 6 are remarkably flat, consistent with the flatter response of the fed funds rate to target rate declines plotted in Figure 5.

Macro aggregate responses in the longer sample, reported in Figure 7, show little effect of a target rate change on inflation. Estimated effects of rate increases on IP suggest less of a decline in response to rate increases than that estimated for the short sample, with a decline of just over 1 pp after two years (versus 1.5 pps in the shorter sample). The unemployment rate response to a rate increase is also somewhat smaller than that found in the shorter sample, while target rate declines are again estimated to have little effect on any macro aggregate.

Our short "crisis sample" saw no target rate increases, so the analysis of this sample models policy changes as any rate decrease of 25 bps or more. The policy propensity score used for this is a probit function of the combined futures variable, along with inflation and unemployment (the Taylor variables used for the policy model estimated in the longer sample). The resulting marginal effects, reported in Table 3, are normalized to show the impact on the probability the target rate is left unchanged, so the signs align with the ordered estimates reported earlier. As with the ordered score model, column (3) in Table 3 shows *FFF* to be a strong predictor of target rate changes. Though the coefficient on inflation is marginally significant (and wrong-signed) in column (3), overall, inflation and unemployment add little to policy prediction conditional on *FFF*, a finding that emerges by comparing the log likelihood in columns (2) and (3). The *F*-statistic is 1.98 with a p-value of 0.14. Finally, a model with inflation and unemployment alone does not predict target rate decreases as well as a model with only *FFF*, as can be seen in a comparison of likelihoods across columns (1) and (2). In fact, in the short crisis sample, traditional Taylor Rule variables are essentially unrelated to target rate declines. The *F*-statistic between columns (1) and (3) is 14.52 with a p-value of nearly 0, suggesting that the term *FFF* is doing the heavy lifting.

Figure 8 shows the estimated response of the federal funds rate to target rate changes using only the crisis period. The results show large but highly imprecisely estimated declines in the federal funds rate. Part of the increase in magnitude here may be due to the fact that the policy dummy in this case indicates a combination of quarter-point and half-point declines.

Figure 9 plots term rate responses to policy interventions in the crisis sample. The estimates offer some evidence that federal funds rate declines are passed down the yield curve in the form of a pattern reasonably consistent with that found when estimated using more data. Finally, Figure 10 reports the responses of the macro variables. Here too the estimates are imprecise. Although inconclusive, they offer little support for the view that monetary stimulus was effective in the crisis period.

## 6 Summary and Conclusions

We start with the assumption that policy changes are independent of potential outcomes, conditional on observed market-based forecasts of these changes plus a small set of economic predictors. Selection-on-observables is a strong assumption, but a natural starting point. We then consider how best to make use of the selection-on-observables identification condition in a potential outcomes framework. The resulting propensity score weighting estimator captures possibly nonlinear and asymmetric causal responses to an ordered dynamic treatment through a simple reweighting procedure. Our framework focuses modeling effort and specification testing on the process that determines policy decision; the model for outcomes is left unspecified.

Our propensity score weighting estimator for ordered time series treatments is applied to evaluate the effect of monetary policy interventions on macroeconomic outcomes before and during the Great Recession. Results for the pre-recession period suggest an asymmetric response to changes in the fed funds rate target, much as implied by the string metaphor. Our findings suggest that target rate increases reduce inflation, employment and output. At the same time, target decreases appear to have little effect on output or inflation. Perhaps surprisingly, an extension of our analysis to cover the "zero lower bound years" since 2008 leaves these findings essentially unchanged.

What explains the asymmetric response of macro aggregates to monetary policy interventions? An important finding emerging from the analysis reported here is the weak effect of target rate declines on medium and long term bond rates. Because changes in these rates provide a plausible causal channel for policy, the relative unresponsiveness of these bond rates to policy may account for much of the weak impact of target rate reductions on macro aggregates. At the same time, the limited precision of many of our estimates means we can't rule out differential sensitivity of macro aggregates to bond yield declines and hikes. In view of this constellation of findings, suggesting though not establishing a "weak channels" hypothesis for asymmetric policy effects, in future work, we plan to use the econometric methods developed here to analyze the effects of the Fed's recent large scale asset purchases. These interventions were designed to reduce the long end of the yield curve in the face of exceptionally low short-term rates.

## A Asset Price Based Policy Predictions

Our formulation of the propensity score is based on Piazzesi’s (2005) term structure model. Piazzesi (2005) provides an explicit parametric framework that links Fed-policy actions to the yield curve. Her model consists of a monetary policy rule  $p_\zeta^j(\zeta_t, \psi)$ , the probability of choosing  $D_t = d_j$ , conditional on information  $\zeta_t$ . Monetary policy then affects the state price density  $M_t$  and consequently, through no-arbitrage arguments, the yield curve. The key insight of Piazzesi’s model that is relevant here is the fact that asset prices, in particular bond yields and related derivatives, depend on the same state vector  $\zeta_t$  as the policy function  $p_\zeta^j(\zeta_t, \psi)$ . Under the additional assumption that the pricing functions are invertible, we can recover  $\zeta_t$  with a vector of asset prices. This is particularly appealing because some of the elements in  $\zeta_t$  may not be directly observable.

We rely on a no-arbitrage pricing relationship between the price of zero coupon bonds as well as a federal funds futures contract and the state vector  $\zeta_t$ . Let the state price density  $M_t$  be such that the price at time  $t$  of a random payoff  $V$  at time  $T$  is  $E[M_T V | \zeta_t] / M_t = E_T^Q[V]$  where  $E_T^Q[V]$  is the expectation operator with respect to the risk neutral measure. Harrison and Kreps (1979) show that the existence of a state price density is essentially equivalent to the existence of an equilibrium price system, something we impose as an assumption. Using the risk neutral measure, random payoffs at various maturities are priced as

$$P_{V_T}(t, T | \zeta_t) = E_T^Q[V | \zeta_t]. \quad (12)$$

We assume that the relationship between the state vector  $\zeta_t$  and (12) is invertible, an assumption that is satisfied for example in affine models. Let  $z_t = (P_{V_1}(t, T_1 | \zeta_t), \dots, P_{V_q}(t, T_q | \zeta_t))$  be a vector of observed price data with maturities  $T_1, \dots, T_q$  and assume that the pricing function has an inverse  $g$  such that

$$\zeta_t = g(z_t, t, T_1, \dots, T_q). \quad (13)$$

The technique of inverting the yield curve to elicit unobservable state variables is well established in the financial econometrics literature – see Söderlind and Svensson (1997), Singleton (2001) or Piazzesi (2005) for examples. Our empirical model for the propensity score is related to the policy function by

$$p^j(z_t, \psi) := p_\zeta^j(g(z_t, t, T_1, \dots, T_q), \psi). \quad (14)$$

Federal funds futures maturing shortly after FOMC announcements are probably good candidates for  $z_t$ . The reason is that there is a direct link between their expected future cash flow and changes in the federal funds target rate. We focus on the case where no FOMC meeting is announced. Because macroeconomic data is released at different days throughout the month and because we are interested in good predictors of expected Fed policy for the entire month  $m + 1$ , we concentrate our attention on the futures price on the last day of the prior month  $m$ , that is,  $\bar{t}_m$ . Let  $f_{\bar{t}_m, m}^1$  denote the price of a one-month ahead contract

traded at the last day of month  $m$  and let  $\kappa_m$  be the total number of days in month  $m$ . On any given day  $t_j$  in month  $m + 1$ , let  $r_{t_j}$  be the effective federal funds rate at the close of the market.<sup>8</sup> The payoff for a buyer of a fed funds futures contract is the difference between the futures rate  $f_{t,m}^1$  and the average fed funds rate over month  $m + 1$ ,

$$r_{t_{m+1}}^a = \frac{1}{\kappa_{m+1}} \sum_{j=1}^{\kappa_{m+1}} r_{t_j} \quad (15)$$

with the payoff cash settled the day after expiration of the futures contract (see Piazzesi and Swanson, 2008, p. 679). Pricing equation (12) and (15) imply that the spread between a funds future  $f_{\bar{t}_m,m}^1$  and the prevailing target rate  $\bar{r}_{\bar{t}_m}$  at the last day  $\bar{t}_m$  of month  $m$  is

$$f_{\bar{t}_m,m}^1 - \bar{r}_{\bar{t}_m} = \frac{E^Q [r_{m+1}^a - \bar{r}_{\bar{t}_m} | \zeta_{\bar{t}_m}]}{P_1(\bar{t}_m, \bar{t}_{m+1} | \zeta_{\bar{t}_m})} \quad (16)$$

where  $\bar{t}_{m+1}$  denotes the end of month  $m + 1$  and  $P_1(\bar{t}_m, \bar{t}_{m+1} | \zeta_{\bar{t}_m})$  is the  $\bar{t}_m$  price of a zero coupon bond maturing at  $\bar{t}_{m+1}$ . Note that  $f_{\bar{t}_m,m}^1$  reflects both uncertainty about whether and when a target rate change will occur in month  $m + 1$  and more general uncertainty about the economy captured by the pricing kernel  $M_t$ . Equation (16) shows that the futures-target rate spread is the best risk adjusted predictor of a target rate change during the coming month.<sup>9</sup> Whether (16) can be inverted to recover  $\zeta_{\bar{t}_m}$  as in (13) depends on the dimension of  $\zeta_t$  as well as the exact functional form of the conditional expectations. In the absence of an explicit pricing model, which would require a more parametric framework than we are willing to entertain, it is ultimately an empirical question whether controlling for  $f_{\bar{t}_m,m}^1 - \bar{r}_{\bar{t}_m}$  in the propensity score is sufficient.

## B Implementation

### B.1 Data

- a) Federal Funds Futures: CBOT prices of current month and next month Federal Funds Futures contract at market close, cry out market Monday - Friday. Source: Bloomberg.
- b) Treasury Bond yields: Daily data from the Federal Reserve Board of Governors. Federal Reserve Statistical Release H.15. Daily observations for Market yields on U.S. Treasury securities at 1-month constant maturity, quoted on investment basis at 3 month and 2, and 10 year maturities. Respective data-codes are ‘RIFLGFCM03’, ‘RIFLGFCY02’, ‘RIFLGFCY10’. Federal funds rate: Daily data based on weighted average of brokered trades. Data-code ‘RIFSPFF\_N.B’

<sup>8</sup>The rate  $r_{t_j}$  is published by the Federal Reserve Bank of New York. It is the volume weighted daily average of trades arranged by major brokers (source: <http://www.newyorkfed.org/markets/omo/dmm/fedfundsdata.cfm>). Using the model implied rate at the end of the day thus is a slight simplification.

<sup>9</sup>In other words, it minimizes the squared prediction error amongst all predictors based on  $\zeta_{\bar{t}_m}$  of  $r_{t_{m+1}}^a$  and under the risk neutral discounted measure.

- c) Macro Data, all monthly frequency. Source: Federal Reserve Bank of St. Louis, Fred. (i) Consumer Price Index of all urban consumers: all items (PCU) seasonally adjusted, Source: Bureau of Labor Statistics; (ii) Industrial Production Index seasonally adjusted (IP), Source: Board of Governors of the Federal Reserve System; (iii) Federal Funds effective Rate in percent per annum (FFED), monthly average, Source: Federal Reserve Statistical Release H.15; (iv) Civilian Unemployment Rate, seasonally adjusted (UNRATE), Source: Bureau of Labor Statistics.

## B.2 Standard Errors

The estimator  $\hat{\alpha}$  defined in (11) is computed in two steps. First, the unconstrained estimate  $\hat{\theta} = T^{-1} \sum_{t=1}^T \hat{h}_t$  is obtained. Let  $\Omega_\theta$  be the asymptotic covariance matrix of  $\hat{\theta}$ . Assume that  $\hat{\Omega}_\theta$  is a consistent estimator of  $\Omega_\theta$ . Now, the estimator  $\hat{\alpha}$  is given as

$$\hat{\alpha} = \arg \min_{\alpha} \left( T^{-1} \sum_{t=1}^T \hat{h}_t - \theta(\alpha) \right)' \hat{\Omega}_\theta^{-1} \left( T^{-1} \sum_{t=1}^T \hat{h}_t - \theta(\alpha) \right).$$

Replacing  $\Omega_\theta$  with  $\hat{\Omega}_\theta$  has no effect on the first order asymptotic distribution of  $\hat{\alpha}$  under our assumptions. On the other hand, the limiting distributions of  $\hat{\theta}$ ,  $\theta(\hat{\alpha})$  and  $\hat{\alpha}$ , derived in Theorems 1 and 2 in Appendix B.5, depend on a preliminary estimate of the the propensity score with parameters  $\psi$ . Assume  $\hat{\psi}$  is the maximum likelihood estimator with representation

$$\tau^{1/2} (\hat{\psi} - \psi) = \Omega_\psi^{-1} T^{-1/2} \sum_{t=1}^T l(D_t, z_t, \psi_0) + o_p(1) \quad (17)$$

where  $\Omega_\psi = E[l(D_t, z_t, \psi_0)l(D_t, z_t, \psi_0)']$  and the function

$$l(D_t, z_t, \psi) = \sum_{j=0}^J \frac{1\{D_t = d_j\}}{p_t^j(z_t, \psi)} \frac{\partial p_t^{d_j}(z_t, \psi)}{\partial \psi}$$

is the score of the maximum likelihood estimator. The representation in (17) is used to expand  $\hat{h}_t$  around  $\psi_0$  leading to  $\hat{h}_t = v_t(\psi_0) + O_p(T^{-1})$  where  $v_t(\psi_0) = h_t(\psi_0) - \theta_0 + \dot{h}(\psi_0)\Omega_\psi^{-1}l(D_t, z_t, \psi_0)$  and  $\dot{h}(\psi_0) = E[\partial h_t(\psi_0)/\partial \psi']$ . The covariance matrix  $\Omega_\theta$  is the typical spectrum at frequency zero matrix of  $v_t(\psi_0)$  found in the HAC-standard error literature (see Newey and West (1994)) and is given by

$$\Omega_\theta = \sum_{i=-\infty}^{\infty} E[v_t(\psi_0)v_{t-i}(\psi_0)'] \quad (18)$$

The formula for  $\Omega_\theta$  takes into account that the ‘observations’  $\hat{h}_t$  used to compute the sample averages are based on estimated, rather than observed data. Confidence intervals for  $\theta$  can be constructed from  $\Omega_\theta$ .

We estimate  $\Omega_\theta$  from the sample averages

$$\hat{h}(\psi_0) = T^{-1} \sum_{t=1}^T \partial h_t(\hat{\psi}) / \partial \psi', \hat{\Omega}_\psi = T^{-1} \sum_{t=1}^T l(D_t, z_t, \hat{\psi}) l(D_t, z_t, \hat{\psi})'$$

and by letting  $v_t(\hat{\psi}) = h_t(\hat{\psi}) - \hat{\theta} + \hat{h}(\hat{\psi}) \hat{\Omega}_\psi^{-1} l(D_t, z_t, \hat{\psi})$ . As in Newey and West (1994), we use the Bartlett kernel with prewhitening and a data-dependent plug in estimator to obtain the necessary bandwidth parameter.

The Newey and West procedure is implemented as follows. Prewhitening is achieved by fitting a VAR(1) model to  $v_t(\hat{\psi})$ . For this purpose define the autoregressive parameter estimate

$$\hat{A} = \sum_{t=2}^T v_t(\hat{\psi}) v_{t-1}(\hat{\psi})' \left( \sum_{t=2}^T v_{t-1}(\hat{\psi}) v_{t-1}(\hat{\psi})' \right)^{-1}$$

and let  $\hat{r}_t(\hat{\psi}) = v_t(\hat{\psi}) - \hat{A} v_{t-1}(\hat{\psi})$ . Then define  $\hat{\Omega}_{\theta,j} = T^{-1} \sum_{t=j+1}^T \hat{r}_t(\hat{\psi}) \hat{r}_{t-j}(\hat{\psi})'$  for  $j \geq 0$  and  $\hat{\Omega}_{\theta,j} = \hat{\Omega}'_{\theta,-j}$  for  $j < 0$ . Let  $\mathbf{1} = [1, \dots, 1]'$  be an  $r$ -dimensional vector where  $r$  is the dimension of  $\theta$ . Define  $\hat{\sigma}_j = \mathbf{1}' \hat{\Omega}_{\theta,j} \mathbf{1}$ ,  $\hat{s}^{(q)} = \sum_{j=-n}^n |j|^q \hat{\sigma}_j$  and  $\hat{\gamma} = c_\gamma (\hat{s}^{(1)}/\hat{s}^{(0)})^{2/3}$  where<sup>10</sup>  $c_\gamma = 1.1447$  and  $n = \lfloor 3(T/100)^{2/9} \rfloor$  where  $\lfloor \cdot \rfloor$  denotes the integer part of a real number. Set the bandwidth parameter to  $\hat{B} = \lfloor \hat{\gamma} T^{1/3} \rfloor$ .

The estimator for  $\Omega_\theta$  is now defined as

$$\hat{\Omega}_\theta = (I_r - \hat{A})^{-1} \left( \hat{\Omega}_{\theta,0} + \sum_{j=1}^{\hat{B}} \left( 1 - \frac{j}{\hat{B}+1} \right) (\hat{\Omega}_{\theta,j} + \hat{\Omega}'_{\theta,j}) \right) (I_r - \hat{A})^{-1}$$

### B.3 Specification Tests

The specification tests are based on the following fact. If  $w_t$  is a vector of elements of  $z_t$  or  $\chi_{t-1}$ , then correct specification of the propensity score implies that

$$E[\mathbf{1}\{D_t = j\} - p^j(z_t, \psi_0) | w_t] = 0 \text{ for all } j = 1, \dots, J.$$

All  $J$  conditional moment restrictions, or a subset of them, can be summarized into a vector. Let  $D_{j,t} = \mathbf{1}\{D_t = d_j\}$ ,  $\mathcal{D}_t = (D_{j_1,t}, \dots, D_{j_k,t})$  and  $p(z_t, \psi) = (p^{j_1}(z_t, \psi), \dots, p^{j_k}(z_t, \psi))$ . Set  $k \leq J$  and  $1 \leq j_1 < \dots < j_k \leq J$ . In our case, we use this setup to focus on  $d_j = \{-.25, 0, .25\}$ . Then,  $E[\mathcal{D}_t - p(z_t, \psi_0) | w_t] = 0$  must hold. To test this condition, consider the statistic  $m(\mathcal{D}_t, z_t, w_t, \psi) = (\mathcal{D}_t - p(z_t, \psi)) \otimes w_t$  for which the unconditional moment restriction  $E[m(\mathcal{D}_t, z_t, w_t, \psi_0)] = 0$  holds. A test of this restriction is based on the limiting distribution of  $T^{-1/2} \sum_{t=1}^T m(\mathcal{D}_t, z_t, w_t, \psi_0)$  under the null-hypothesis of correct specification of  $p(z_t, \psi_0)$ .

The testing problem is complicated by the fact that  $\psi_0$  is unknown and needs to be estimated. This

<sup>10</sup>See Newey and West (1994, Tables I and II).

affects the limiting distribution of the test statistic. Define  $\hat{p}(\psi) = T^{-1} \sum_{t=1}^T \partial p(z_t, \hat{\psi}) / \partial \psi' \otimes w_t$  and

$$\hat{\vartheta}_t = \hat{m}_t - \hat{p}(\psi) \hat{\Omega}_\psi^{-1} l(D_t, z_t, \hat{\psi})$$

where the second term in  $\hat{\vartheta}_t$  corrects  $\hat{m}_t$  for the effects of parameter estimation. A key insight is that under the null-hypothesis,  $\hat{\vartheta}_t$  is approximately a martingale difference sequence. This feature significantly simplifies estimation of the asymptotic variance normalizing the test. Then, letting  $\bar{m} = T^{-1} \sum_{t=1}^T \hat{m}_t$  and  $\hat{V} = T^{-1} \sum_{t=1}^T \hat{\vartheta}_t \hat{\vartheta}_t'$  leads to the test statistic

$$T \bar{m}' \hat{V}^{-1} \bar{m} \rightarrow_d \chi_{(J)}^2 \quad (19)$$

under the null hypothesis that  $E[\mathbf{1}\{D_t = j\} | z_t] = p^j(z_t, \psi_0)$ . The limiting distribution in (19) is established in Appendix B.5.

## B.4 Regularity Conditions

Let  $\chi_t = [y_t', x_t', D_t]'$  be the vector of observations. Assume that  $\{\chi_t\}_{t=1}^\infty$  is strictly stationary with values in the measurable space  $(\mathbb{R}^r, \mathcal{B}^r)$  where  $\mathcal{B}^r$  is the Borel  $\sigma$ -field on  $\mathbb{R}^r$  and  $r$  is fixed with  $2 \leq r < \infty$ . Let  $\mathcal{A}_1^l = \sigma(\chi_1, \dots, \chi_l)$  be the sigma field generated by  $\chi_1, \dots, \chi_l$ . The sequence  $\chi_t$  is  $\beta$ -mixing or absolutely regular if

$$\beta_m = \sup_{l \geq 1} E \left[ \sup_{A \in \mathcal{A}_{l+m}^\infty} \left| \Pr(A | \mathcal{A}_1^l) - P(A) \right| \right] \rightarrow 0 \text{ as } m \rightarrow \infty.$$

**Condition 2** Let  $\chi_t$  be a stationary, absolutely regular sequence such that for some  $2 < p < \infty$  the  $\beta$ -mixing coefficient of  $\chi_t$  satisfies  $\beta_m \leq cm^{-\frac{1+p}{p-2/p}}$  for some bounded constant  $c > 0$ .

Condition 2 implies that  $\sum_{m=1}^\infty \beta_m^{1-1/p} < \infty$  as required for Corollary 3.9 of McLeish (1975a). In addition,  $\beta_m$  satisfies (2.6) of McLeish (1975b) required for a strong law of large numbers. Using Corollary A.2 of Hall and Heyde (1980), it also follows that  $\sum_{i=m}^\infty |m|^p \|E[v_t(\psi_0) v_{t-m}(\psi_0)']\| < \infty$  as required by Assumption 2 of Newey and West (1994). Also note that  $p > 2$  is sufficient to satisfy Assumption 3 of Newey and West (1994) when the Bartlett kernel is used as suggested here.

The next condition states that the propensity score  $p(z_t, \theta)$  is the correct parametric model for the conditional expectation of  $D_t$  and lists a number of additional regularity conditions.

**Condition 3** Let  $\psi_0 \in \Psi$  where  $\Psi \subset \mathbb{R}^{k_\psi}$  is a compact set and  $k_\psi < \infty$ . Assume that  $E[\mathbf{1}\{D_t = d_j\} | z_t] = p_t^j(z_t, \psi_0)$  and for all  $\psi \neq \psi_0$  it follows  $E[\mathbf{1}\{D_t = d_j\} | z_t] \neq p^j(z_t | \psi)$ . Assume that  $p^j(z_t | \psi)$  is differentiable a.s. for  $\psi \in \{\psi \in \Theta \mid \|\psi - \psi_0\| \leq \delta\} := N_\delta(\psi_0)$  for some  $\delta > 0$ . Let  $N(\psi_0)$  be a compact subset of the union of all neighborhoods  $N_\delta(\psi_0)$  where  $\partial p^j(z_t | \psi) / \partial \psi$ ,  $\partial^2 p^j(z_t | \psi) / \partial \psi_i \partial \psi_j$  exists and assume that  $N(\psi_0)$  is not empty. Assume that for all  $j \in \{0, \dots, J\}$  and some  $\delta_0 > 0$  and any  $\delta > 0$   $\psi, \psi'$  with

$\|\psi - \psi'\| < \delta \leq \delta_0$  there exists a random variable  $B_t$  which is a measurable function of  $D_t, z_t$  and  $Y_{t,L}$  and a constant  $\alpha > 0$  such that for all  $i$

$$\|h_{t,j}(\psi) - h_{t,j}(\psi')\| \leq B_t \|\psi - \psi'\|^\alpha,$$

and

$$\|\partial h_{t,j}(\psi) / \partial \psi - \partial h_{t,j}(\psi') / \partial \psi\| \leq B_t \|\psi - \psi'\|^\alpha \quad (20)$$

$$\|z_t (\delta_{t,j}(\psi) - \delta_{t,j}(\psi'))\| \leq B_t \|\psi - \psi'\|^\alpha \quad (21)$$

and  $\psi, \psi' \in \text{int } N(\psi_0)$ . Let  $h_{t,j,i}(\psi)$  be the  $i$ -th element of  $h_{t,j}(\psi)$ .  $E[B_t^p] < \infty$ , for all  $i$   $E[|h_{t,j,i}(\psi_0)|^p] < \infty$ ,  $E[|\partial h_{t,j,i}(\psi_0) / \partial \psi|^p] < \infty$ .

**Condition 4** Assume that  $\hat{\psi} - \psi_0 = o_p(1)$ ,  $T^{1/2}(\hat{\psi} - \psi_0) = \Omega_\psi^{-1} T^{-1/2} \sum_{t=1}^T l(D_t, z_t, \psi_0) + o_p(1)$ . Let  $p$  be given as in Condition 2 and assume that  $E[|l(D_t, z_t, \psi_0)|^p] < \infty$ .

**Condition 5** Assume that  $\Omega_\psi$  is positive definite for all  $\psi$  in some neighborhood  $N \subset \Psi$  such that  $\psi_0 \in \text{int } N$  and  $0 < \|\Omega_\psi\| < \infty$  for all  $\psi \in N$ . Assume that  $\Omega_\theta$  defined in (18) is positive definite.

Conditions 2, 3 and 4 imply that Assumption 2 of Newey and West is satisfied. The results of their paper thus apply to the estimates of  $\Omega_\theta$  proposed here.

Regularity conditions for the specification tests are given below.

**Condition 6** Let  $N(\psi_0)$  be the set defined in Condition 3. For some random variable  $B_t$  which is a measurable function of  $D_t, z_t$  and  $w_t$  and for which  $E[B_t^p] < \infty$ , it holds that for some  $\varepsilon > 0$  and  $\psi, \psi'$  with  $\|\psi - \psi'\| < \delta \leq \delta_0$  and  $\psi, \psi' \in \text{int } N(\psi_0)$  that

$$i) E[\|\vartheta_t\|^{p+\varepsilon}] < \infty, E[\|\partial p(z_t, \psi_0) / \partial \psi' \otimes w_t\|^{p+\varepsilon}] < \infty, E[\|l(D_t, z_t, \psi_0)\|^{p+\varepsilon}] < \infty$$

$$ii) \|l(D_t, z_t, \psi) - l(D_t, z_t, \psi')\| \leq B_t \|\psi - \psi'\|^\alpha,$$

$$iii) \|(\partial p(z_t, \psi) / \partial \psi' - \partial p(z_t, \psi') / \partial \psi') \otimes w_t\| \leq B_t \|\psi - \psi'\|^\alpha.$$

## B.5 Proofs

**Theorem 1** Let  $\hat{\theta}$  be defined in (9) and assume that Conditions 1, 2, 3, 4, and 5 hold. Then,

$$T^{1/2}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Omega_\theta)$$

where  $\Omega_\theta$  is defined in (18).

**Proof.** Let  $Z = (z_1, \dots, z_T)'$ ,  $Y_L = (Y_{1,L}, \dots, Y_{T,L})'$  and  $\delta_j(\hat{\psi}) = (\delta_{1,j}(\hat{\psi}), \dots, \delta_{T,j}(\hat{\psi}))'$ . Define the population projection  $\pi_y$  as  $\pi_y = \arg \min_b E[\|Y_{t,L} - bz_t\|^2]$  and sample analog  $\hat{\pi}_y = Y_L' Z (Z' Z)^{-1}$ . Recall that  $\hat{h}_t = Y_{t,L} (\delta_{t,j}(\hat{\psi}) - \hat{\delta}_{t,j})$  where  $\hat{\delta}_{t,j} = z_t'(Z' Z)^{-1} Z' \delta_j(\hat{\psi})$  and let  $h_t(\psi_0) = (Y_{t,L} - \pi_y z_t) \delta_{t,j}(\psi_0)$ .

By the Mean Value Theorem we obtain

$$\begin{aligned}
T^{1/2} (\hat{\theta} - \theta_0) &= T^{-1/2} \sum_{t=1}^T \hat{h}_t - \theta_0 \\
&= T^{-1/2} \sum_{t=1}^T (Y_{t,L} - \pi_y z_t) \delta_{t,j} (\hat{\psi}) - \theta_0 + (\pi_y - \hat{\pi}_y) T^{-1/2} \sum_{t=1}^T z_t \delta_{t,j} (\hat{\psi}) \\
&= T^{-1/2} \sum_{t=1}^T h_t(\psi_0) - \theta_0 + T^{-1} \sum_{t=1}^T \partial h_t(\psi_0) / \partial \psi' (T/\tau)^{1/2} T^{1/2} (\hat{\psi} - \psi_0) \\
&\quad + T^{-1} \sum_{t=1}^T (\partial h_t(\check{\psi}) / \partial \psi' - \partial h_t(\psi_0) / \partial \psi') T^{1/2} (\hat{\psi} - \psi_0) \\
&\quad + (\pi_y - \hat{\pi}_y) T^{-1/2} \sum_{t=1}^T z_t \delta_{t,j} (\hat{\psi})
\end{aligned} \tag{22}$$

where  $\|\check{\psi} - \psi_0\| \leq \|\hat{\psi} - \psi_0\|$  and  $\partial h_t(\psi) / \partial \psi' = [\partial h_{t,1}(\psi) / \partial \psi', \dots, \partial h_{t,J}(\psi) / \partial \psi']$  with

$$\partial h_{t,j}(\psi) / \partial \psi = Y_{t,L} \left( -\frac{D_{t,j}}{p^j(z_t, \psi)^2} \frac{\partial p^j(z_t, \psi)}{\partial \psi} + \frac{D_{t,0}}{p^0(z_t, \psi)^2} \frac{\partial p^0(z_t, \psi)}{\partial \psi} \right). \tag{23}$$

By (20) it follows that for  $\delta_0$  given in Condition 3 and any  $\delta$  such that  $\delta_0 > \delta > 0$ ,

$$\begin{aligned}
&P \left( \left\| T^{-1} \sum_{t=1}^T (\partial h_{t,j}(\check{\psi}) / \partial \psi' - \partial h_{t,j}(\psi_0) / \partial \psi') \right\| > \eta \right) \\
&\leq P \left( \sup_{\|\psi - \psi_0\| \leq \delta} \left\| T^{-1} \sum_{t=1}^T (\partial h_{t,j}(\psi) / \partial \psi' - \partial h_{t,j}(\psi_0) / \partial \psi') \right\| > \eta, \|\check{\psi} - \psi_0\| < \delta \right) + P(\|\check{\psi} - \psi_0\| \geq \delta) \\
&= \frac{E[|B_t|^p] \delta^{p\alpha}}{\eta^p} + P(\|\check{\psi} - \psi_0\| \geq \delta)
\end{aligned} \tag{24}$$

where both terms can be made arbitrarily small by choosing  $\eta = \sqrt{\delta}$  and  $\delta > 0$  for  $T$  large enough by using Conditions 4 and 3. By McLeish (1975b, Theorem 2.10)  $T^{-1} \sum_{t=1}^T \partial h_{t,j}(\psi_0) / \partial \psi' \xrightarrow{p} \dot{h}(\psi_0)$  where we defined  $E[\partial h_{t,j}(\psi_0) / \partial \psi'] = \dot{h}_j(\psi_0)$ . This implies that the third term in (22) is  $o_p(1)$ .

For the last term in (22) note that  $(\pi_y - \hat{\pi}_y) = O_p(T^{-1/2})$  by McLeish (1975b, Theorem 2.10), Corollary 3.9 of McLeish (1975a) and standard arguments for linear regressions. Now consider

$$\begin{aligned}
&(\pi_y - \hat{\pi}_y) T^{-1/2} \sum_{t=1}^T z_t \delta_{t,j} (\hat{\psi}) \\
&= T^{1/2} (\pi_y - \hat{\pi}_y) T^{-1} \sum_{t=1}^T z_t \delta_{t,j} (\psi_0) \\
&\quad + T^{1/2} (\pi_y - \hat{\pi}_y) T^{-1} \sum_{t=1}^T z_t \left( \delta_{t,j} (\hat{\psi}) - \delta_{t,j} (\psi_0) \right).
\end{aligned} \tag{25}$$

The first term in (25) is  $o_p(1)$  because from  $E[z_t \delta_{t,j}(\psi_0)] = 0$  it follows that

$$T^{-1} \sum_{t=1}^T z_t \delta_{t,j}(\psi_0) = o_p(1). \tag{26}$$

For the second term in (25) use Condition 3 to show that

$$T^{-1} \sum_{t=1}^T z_t \left( \delta_{t,j}(\psi_0) - \delta_{t,j}(\hat{\psi}) \right) = o_p(1) \quad (27)$$

by arguments similar to those in (24). Then, (26) and (27) establish that (25) is  $o_p(1)$

It then follows from (25) and (27) that (22) is

$$\begin{aligned} & T^{-1/2} \sum_{t=1}^T h_{t,j}(\psi_0) - \theta_0 \\ & + T^{-1} \sum_{t=1}^T \partial h_{t,j}(\psi_0) / \partial \psi' T^{1/2} (\hat{\psi} - \psi_0) + o_p(1) \\ = & T^{-1/2} \sum_{t=1}^T \left[ h_{t,j}(\psi_0) - \theta_0 + \dot{h}_j(\psi_0) \Omega_\psi^{-1} l(D_t, z_t, \psi_0) \right] + o_p(1). \end{aligned}$$

Let

$$v_t(\psi_0) = h_t(\psi_0) - \theta + \dot{h}_j(\psi_0) \Omega_\psi^{-1} l(D_t, z_t, \psi_0)$$

and  $v_{t,j}(\psi_0)$  is the  $j$ -th element of  $v_t(\psi_0)$ . Note that  $v_{t,j}(\psi_0)$  is  $\beta$ -mixing with  $E[v_{t,j}(\psi_0)] = 0$ . Then it follows that

$$T^{-1} E \left[ \sum_{t=1}^T \sum_{s=1}^T v_t(\psi_0) v_s(\psi_0)' \right] \quad (28)$$

$$= \sum_{j=-T+1}^{T-1} \left( 1 - \frac{|j|}{T} \right) E \left[ v_1(\psi_0) v_{1-j}(\psi_0)' \right] \rightarrow \Omega_\theta \quad (29)$$

by stationarity of  $v_t(\psi_0)$  and the Toeplitz lemma. Fix  $\lambda \in \mathbb{R}^k$  with  $\|\lambda\| = 1$  and let  $S_T = T^{-1/2} \sum_{t=1}^T \lambda' v_t$ . Then,  $E[S_T^2] \rightarrow \lambda' \Omega_\theta \lambda > 0$  by (28) and Condition 5. In addition

$$E \left[ |\lambda' v_t|^p \right] \leq E \left[ \left( \sum_{l=1}^k |\lambda_l| |\tilde{v}_{t,l}| \right)^p \right] \leq \left( \sum_{l=1}^k |\lambda_l|^{\frac{p}{p-1}} \right)^{p-1} E \left[ \sum_{l=1}^k |\tilde{v}_{t,l}|^p \right]$$

by Hölder's inequality (Magnus and Neudecker, 1988, p.220) and where  $\tilde{v}_{t,l}$  is the  $l$ -th element of  $v_t$ . Since  $p/p - 1 \leq 2$  and  $\|\lambda\| = 1$  it follows that  $\sum_{l=1}^k |\lambda_l|^{\frac{p}{p-1}} < 1$ . Denote by  $h_{t,j}(\psi_0)$  and  $\theta_{(j)}$  the  $j$ -th element of  $h_t(\psi_0)$  and  $\theta$  respectively and by  $\dot{h}_j(\psi_0)$  the  $j$ -th row of  $\dot{h}(\psi_0)$ . Then,

$$\begin{aligned} E[|\tilde{v}_{t,j}|^p] & \leq E \left[ \left( |h_{t,j}(\psi_0)| + |\theta_{(j)}| + \left\| \dot{h}_j(\psi_0) \right\| \left\| \Omega_\psi^{-1} \right\| \|l(D_t, z_t, \psi_0)\| \right)^p \right] \\ & \leq 3^{p-1} \left( E[|h_{t,j}(\psi_0)|^p] + |\theta_j|^p + \left\| \dot{h}_j(\psi_0) \right\|^p \left\| \Omega_\psi^{-1} \right\|^p \|l(D_t, z_t, \psi_0)\|^p \right) \end{aligned}$$

again by Hölder's inequality. It follows that  $|\theta_{(j)}|^p \leq E[|h_{t,j}(\psi_0)|^p]$  by Jensen's inequality and  $\left\| \Omega_\psi^{-1} \right\|^p < \infty$  by Condition 5. Similarly,  $E[\|l(D_t, z_t, \psi_0)\|^p] < \infty$  by Condition 4 and  $\left\| \dot{h}_j(\psi_0) \right\|^p \leq E[|\partial h_{t,j}(\psi_0) / \partial \psi|^p] < \infty$  by Condition 3. By Condition 3  $E[|h_{t,j}(\psi_0)|^p] < \infty$  such that  $E[|\tilde{v}_{t,j}|^p] < \infty$ . These arguments together with Condition 2 show that all the conditions of Corollary 3.9 of McLeish (1975a) are satisfied.

Thus,  $S_T \rightarrow_d N(0, \lambda' \Omega_\theta \lambda)$ . The result now follows from the Cramer-Wold theorem. ■

**Theorem 2** *Let  $\hat{\alpha}$  be defined in (11) and assume that Conditions 1, 2, 3, 4, and 5 hold. Let  $\dot{\theta}(\alpha) = \partial\theta(\alpha)/\partial\alpha'$  and assume that  $\dot{\theta}(\alpha_0) \equiv A$  has full column rank. In addition, assume that  $\theta_0 - \theta(\alpha) = 0$  if and only if  $\alpha = \alpha_0$  where  $\alpha_0 \in \mathcal{A} \subset \mathbb{R}^d$  and  $\mathcal{A}$  is compact. Then,*

$$T^{1/2}(\hat{\alpha} - \alpha_0) \xrightarrow{d} N\left(0, (A' \Omega_\theta^{-1} A)^{-1}\right)$$

where  $\Omega_\theta$  is defined in (18). For any  $\lambda \in \mathbb{R}^d$  such that  $A'\lambda \neq 0$  it follows that

$$T^{1/2}\lambda'(\theta(\hat{\alpha}) - \theta(\alpha_0)) \xrightarrow{d} N\left(0, \lambda' A (A' \Omega_\theta^{-1} A)^{-1} A' \lambda\right).$$

**Proof.** Let

$$Q_T(\alpha) = \left(T^{-1} \sum_{t=1}^T \hat{h}_t - \theta(\alpha)\right)' \hat{\Omega}_\theta^{-1} \left(T^{-1} \sum_{t=1}^T \hat{h}_t - \theta(\alpha)\right).$$

By the arguments in the proof of Theorem 1 it follows that  $T^{-1} \sum_{t=1}^T \hat{h}_t \rightarrow_p \theta$ . Thus, uniformly in  $\alpha \in \mathcal{A}$ ,

$$Q_T(\alpha) \rightarrow_p Q(\alpha) = (\theta - \theta(\alpha))' \Omega_\theta^{-1} (\theta - \theta(\alpha)).$$

Because  $\Omega_\theta^{-1}$  is positive definite by Condition 5 and  $\theta - \theta(\alpha) = 0$  if and only if  $\alpha = \alpha_0$  it follows that  $Q(\alpha) \geq 0$  and  $Q(\alpha) = 0$  if and only if  $\alpha = \alpha_0$ . By standard arguments it follows that  $\hat{\alpha} \rightarrow_p \alpha_0$ . We use a mean value expansion of the first order condition

$$\left(\frac{\partial\theta(\hat{\alpha})}{\partial\alpha'}\right)' \hat{\Omega}_\theta^{-1} \left(T^{-1} \sum_{t=1}^T \hat{h}_t - \theta(\hat{\alpha})\right) = 0$$

such that

$$0 = \left(\frac{\partial\theta(\hat{\alpha})}{\partial\alpha'}\right)' \hat{\Omega}_\theta^{-1} \sqrt{T} \left(T^{-1} \sum_{t=1}^T \hat{h}_t - \theta(\alpha_0)\right) + \left(\frac{\partial\theta(\hat{\alpha})}{\partial\alpha'}\right)' \hat{\Omega}_\theta^{-1} \left(\frac{\partial\theta(\tilde{\alpha})}{\partial\alpha'}\right) \sqrt{T} (\hat{\alpha} - \alpha_0) \quad (30)$$

where  $\|\tilde{\alpha} - \alpha_0\| \leq \|\hat{\alpha} - \alpha_0\| \rightarrow_p 0$ . Then, by the continuous mapping theorem, it follows that  $\partial\theta(\hat{\alpha})/\partial\alpha' \rightarrow_p \dot{\theta}(\alpha_0)$  and  $\partial\theta(\tilde{\alpha})/\partial\alpha' \rightarrow_p \dot{\theta}(\alpha_0)$ . By the same arguments as in the proof of Theorem 1 it follows that  $T^{-1/2} \sum_{t=1}^T (\hat{h}_t - \theta(\alpha_0)) \rightarrow_d N(0, \Omega_\theta)$ . The result then follows directly from rearranging (30) and applying the continuous mapping theorem.

To prove the second part of the theorem use a mean value expansion around  $\alpha_0$

$$T^{1/2}\lambda'(\theta(\hat{\alpha}) - \theta(\alpha_0)) = \lambda' \partial\theta(\tilde{\alpha})/\partial\alpha' T^{1/2}(\hat{\alpha} - \alpha_0) \rightarrow_d \lambda' \dot{\theta}(\alpha_0) \xi$$

where  $\xi \sim N\left(0, (A' \Omega_\theta^{-1} A)^{-1}\right)$ . The result then follows from the continuous mapping theorem, noting that  $\lambda' A (A' \Omega_\theta^{-1} A)^{-1} A' \lambda$  is non-singular. ■

The following theorem establishes the limiting distribution of the test statistic in (19).

**Theorem 3** Assume that Conditions 2, 3, 4, 5 and 6 hold. Let  $V_t = \vartheta_t \vartheta_t' - V$  where  $\vartheta_t$  is defined in (31). Assume that for any element  $V_{t,ij}$  of  $V_t$ ,  $E[|V_{t,ij}|^p] < \infty$ . Then,

$$T\bar{m}'\hat{V}^{-1}\bar{m} \rightarrow_d \chi_{(J)}^2$$

**Proof.** Let  $P(z_t, \psi) = \text{diag}(p(z_t, \psi))$ ,  $h(z_t, \psi) = (\text{diag}(p(z_t, \psi)) - p(z_t, \psi)p(z_t, \psi)')$ . Simple algebra then shows that

$$(\mathcal{D}_t - p(z_t, \psi))' h(z_t, \psi)^{-1} \partial p(D_t|z_t, \psi) / \partial \psi' = \partial \ell(\mathcal{D}_t, z_t, \psi) / \partial \psi' = l(D_t, z_t, \psi_0)'$$

where  $\ell(\mathcal{D}_t, z_t, \psi) = \sum_{j=0}^J D_{j,t} \log p^j(z_t, \psi)$  is the log likelihood of the multinomial distribution and  $D_{j,t} = \mathbf{1}\{D_t = d_j\}$ .

Recall  $\hat{m}_t = (\mathcal{D}_t - p(z_t, \hat{\psi})) \otimes w_t$  such that for  $m_{t,0} = m(\mathcal{D}_t, z_t, w_t, \psi_0)$ ,

$$\hat{m}_t = m_{t,0} - (\partial p(z_t, \hat{\psi}) / \partial \psi' \otimes w_t) (\hat{\psi} - \psi_0)$$

with  $\|\hat{\psi} - \psi_0\| \leq \|\hat{\psi} - \psi_0\|$ . Using (17) as well as Condition 4 and setting  $\hat{p}(\hat{\psi}) = T^{-1} \sum_{t=1}^T \partial p(z_t, \hat{\psi}) / \partial \psi' \otimes w_t$  we obtain

$$\begin{aligned} T^{-1/2} \sum_{t=1}^T \hat{m}_t &= T^{-1/2} \sum_{t=1}^T m_{t,0} - \hat{p}(\hat{\psi}) \Omega_{\hat{\psi}}^{-1} T^{-1/2} \sum_{t=1}^T l(D_t, z_t, \psi_0) + o_p(1). \\ &= T^{-1/2} \sum_{t=1}^T m_{t,0} - \dot{p}(\psi_0) \Omega_{\psi_0}^{-1} \partial p(D_t|z_t, \psi_0)' / \partial \psi h(z_t, \psi_0)^{-1} (\mathcal{D}_t - p(z_t, \psi_0)) + o_p(1) \end{aligned}$$

where the last line follows from  $E[\partial p(z_t, \psi_0) / \partial \psi' \otimes w_t] = \dot{p}(\psi_0)$  and  $\hat{p}(\hat{\psi}) - \dot{p}(\psi_0) = o_p(1)$ . Since

$$\vartheta_t = m_{t,0} - \dot{p}(\psi_0) \Omega_{\psi_0}^{-1} \partial p(D_t|z_t, \psi_0)' / \partial \psi h(z_t, \psi_0)^{-1} (\mathcal{D}_t - p(z_t, \psi_0)) \quad (31)$$

is a martingale difference sequence we consider

$$E[m_{t,0} m_{t,0}'] = E[(P(z_t, \psi_0) - p(z_t, \psi_0)p(z_t, \psi_0)') \otimes w_t w_t'] = \Gamma,$$

$$\Omega_{\psi_0} = E[\partial p(D_t|z_t, \psi_0)' / \partial \psi h(z_t, \psi_0)^{-1} \partial p(D_t|z_t, \psi_0) / \partial \psi']$$

and

$$E[m_{t,0} (\mathcal{D}_t - p(z_t, \psi_0))' h(z_t, \psi_0)^{-1} \partial p(D_t|z_t, \psi_0) / \partial \psi'] = E[\partial p(z_t, \psi_0) / \partial \psi' \otimes w_t] = \dot{p}(\psi_0).$$

By Corollary 3.9 of McLeish (1975a) it follows that

$$T^{-1/2} \sum_{t=1}^T \hat{m}_t = T^{-1/2} \sum_{t=1}^T \vartheta_t + o_p(1) \rightarrow^d N(0, V) \quad (32)$$

where  $V = \Gamma - \dot{p}(\psi_0) \Omega_\psi^{-1} \dot{p}(\psi_0)'$ . A detailed verification of the conditions is omitted but follows the same line of argument as given in the proof of Theorem 1 above. To estimate  $V$ , define

$$\hat{\vartheta}_t = \hat{m}_t - \hat{p}(\hat{\psi}) \hat{\Omega}_\psi^{-1} l(D_t, z_t, \hat{\psi})$$

with

$$\hat{\Omega}_\psi = T^{-1} \sum_{t=1}^T l(D_t, z_t, \hat{\psi}) l(D_t, z_t, \hat{\psi})'$$

Let

$$\hat{V} = T^{-1} \sum_{t=1}^T \hat{\vartheta}_t \hat{\vartheta}_t'$$

By arguments similar to the proof of Theorem 1 it follows that

$$\hat{\Omega}_\psi \rightarrow_p \Omega_\psi \quad (33)$$

and

$$\hat{p}(\hat{\psi}) \rightarrow_p \dot{p}(\psi_0). \quad (34)$$

Next, expand

$$\begin{aligned} \hat{\vartheta}_t &= m_{t,0} - (\partial p(z_t, \check{\psi}) / \partial \psi' \otimes w_t) (\hat{\psi} - \psi_0) \\ &\quad - \left( \hat{p}(\hat{\psi}) \hat{\Omega}_\psi^{-1} - \dot{p}(\psi_0) \Omega_\psi^{-1} \right) l(D_t, z_t, \hat{\psi}) \\ &\quad - \dot{p}(\psi_0) \Omega_\psi^{-1} \left( l(D_t, z_t, \hat{\psi}) - l(D_t, z_t, \psi_0) \right) \\ &\quad - \dot{p}(\psi_0) \Omega_\psi^{-1} l(D_t, z_t, \psi_0) \end{aligned}$$

and recalling  $\vartheta_t = m_{t,0} - \dot{p}(\psi_0) \Omega_\psi^{-1} l(D_t, z_t, \psi_0)$ . Then,

$$\left\| T^{-1} \sum_{t=1}^T \hat{\vartheta}_t \hat{\vartheta}_t' - V \right\| \leq \left\| T^{-1} \sum_{t=1}^T \left( \hat{\vartheta}_t \hat{\vartheta}_t' - \vartheta_t \vartheta_t' \right) \right\| + \left\| T^{-1} \sum_{t=1}^T \vartheta_t \vartheta_t' - V \right\| \quad (35)$$

where the second term on the RHS of (35) is  $o_p(1)$  by Theorem 2.10 of McLeish (1995b). Next, consider

$$T^{-1} \sum_{t=1}^T \left( \hat{\vartheta}_t \hat{\vartheta}_t' - \vartheta_t \vartheta_t' \right) = T^{-1} \sum_{t=1}^T \left( \hat{\vartheta}_t - \vartheta_t \right) \left( \hat{\vartheta}_t - \vartheta_t \right)' + \vartheta_t \left( \hat{\vartheta}_t - \vartheta_t \right)' - \left( \hat{\vartheta}_t - \vartheta_t \right) \vartheta_t' \quad (36)$$

where

$$\begin{aligned}\hat{\vartheta}_t - \vartheta_t &= -(\partial p(z_t, \check{\psi})/\partial \psi' \otimes w_t) (\hat{\psi} - \psi_0) - \left( \hat{p}(\hat{\psi}) \hat{\Omega}_{\hat{\psi}}^{-1} - \dot{p}(\psi_0) \Omega_{\psi}^{-1} \right) l(D_t, z_t, \hat{\psi}) \\ &\quad - \dot{p}(\psi_0) \Omega_{\psi}^{-1} \left( l(D_t, z_t, \hat{\psi}) - l(D_t, z_t, \psi_0) \right).\end{aligned}\quad (37)$$

Thus,

$$\begin{aligned}T^{-1} \sum_{t=1}^T \vartheta_t (\hat{\vartheta}_t - \vartheta_t)' &= -T^{-1} \sum_{t=1}^T \vartheta_t \left( (\partial p(z_t, \check{\psi})/\partial \psi' \otimes w_t) (\hat{\psi} - \psi_0) \right)' \\ &\quad - T^{-1} \sum_{t=1}^T \vartheta_t \left( \left( \hat{p}(\hat{\psi}) \hat{\Omega}_{\hat{\psi}}^{-1} - \dot{p}(\psi_0) \Omega_{\psi}^{-1} \right) l(D_t, z_t, \hat{\psi}) \right)' \\ &\quad - T^{-1} \sum_{t=1}^T \vartheta_t \left( \dot{p}(\psi_0) \Omega_{\psi}^{-1} \left( l(D_t, z_t, \hat{\psi}) - l(D_t, z_t, \psi_0) \right) \right)' \\ &\equiv R_1 + R_2 + R_3.\end{aligned}\quad (38)$$

For  $R_1$  note that

$$\begin{aligned}\|R_1\| &\leq \left\| T^{-1} \sum_{t=1}^T \vartheta_t \partial p(z_t, \psi_0)/\partial \psi' \otimes w_t \right\| \|\hat{\psi} - \psi_0\| \\ &\quad + T^{-1} \sum_{t=1}^T \|\vartheta_t\| \|w_t\| \left\| \partial p(z_t, \psi_0)/\partial \psi' - \partial p(z_t, \check{\psi})/\partial \psi' \right\| \|\hat{\psi} - \psi_0\|\end{aligned}\quad (39)$$

where  $\|\hat{\psi} - \psi_0\| = O_p(T^{-1/2})$  and

$$T^{-1} \sum_{t=1}^T \vartheta_t (\partial p(z_t, \psi_0)/\partial \psi' \otimes w_t) = O_p(1) \quad (40)$$

because

$$E \left[ \left\| \vartheta_t \partial p(z_t, \psi_0)/\partial \psi' \otimes w_t \right\|^{(p+\epsilon)/2} \right] \leq \left( E \left[ \|\vartheta_t\|^{p+\epsilon} \right] E \left[ \left\| \partial p(z_t, \psi_0)/\partial \psi' \otimes w_t \right\|^{p+\epsilon} \right] \right)^{1/2} < \infty$$

by Condition 6 and by Theorem 2.10 of McLeish (1975b).<sup>11</sup> The second term in (39) can be bounded with probability approaching 1 as  $T \rightarrow \infty$ , using Condition 6(iii), and noting that

$$\left\| (\partial p(z_t, \psi)/\partial \psi' - \partial p(z_t, \psi')/\partial \psi') \otimes w_t \right\| = \left\| (\partial p(z_t, \psi)/\partial \psi' - \partial p(z_t, \psi')/\partial \psi') \right\| \|w_t\|,$$

by

$$\begin{aligned}&T^{-1} \sum_{t=1}^T \|\vartheta_t\| \|w_t\| \left\| \partial p(z_t, \psi_0)/\partial \psi' - \partial p(z_t, \check{\psi})/\partial \psi' \right\| \|\hat{\psi} - \psi_0\| \\ &\leq \left\| \hat{\psi} - \psi_0 \right\|^{1+\alpha} T^{-1} \sum_{t=1}^T \|\vartheta_t\| |B_t|\end{aligned}\quad (41)$$

<sup>11</sup>We use McLeish (1975), Equation (2.12) and stationarity to establish Condition (2.11) of Theorem (2.10).

where  $E \left[ \|\vartheta_t\|^{(p+\epsilon)/2} |B_t|^{(p+\epsilon)/2} \right] \leq (E [\|\vartheta_t\|^{p+\epsilon}] E [|B_t|^{p+\epsilon}])^{1/2} < \infty$  by Condition 6. This again implies that

$$T^{-1} \sum_{t=1}^T \|\vartheta_t\| |B_t| = O_p(1) \quad (42)$$

by McLeish (1975b). Now (40) and (41) imply that  $R_1 = o_p(1)$ .

For  $R_2$  note that using Condition 6(ii), w.p.a.1 as  $T \rightarrow \infty$ ,

$$\begin{aligned} \|R_2\| &\leq \left\| \hat{p}(\hat{\psi}) \hat{\Omega}_{\hat{\psi}}^{-1} - \dot{p}(\psi_0) \Omega_{\psi}^{-1} \right\| T^{-1} \sum_{t=1}^T \|\vartheta_t\| \|l(D_t, z_t, \psi_0)\| \\ &\quad + \left\| \hat{p}(\hat{\psi}) \hat{\Omega}_{\hat{\psi}}^{-1} - \dot{p}(\psi_0) \Omega_{\hat{\psi}}^{-1} \right\| T^{-1} \sum_{t=1}^T \|\vartheta_t\| \left\| l(D_t, z_t, \psi_0) - l(D_t, z_t, \hat{\psi}) \right\| \\ &\leq \left\| \hat{p}(\hat{\psi}) \hat{\Omega}_{\hat{\psi}}^{-1} - \dot{p}(\psi_0) \Omega_{\hat{\psi}}^{-1} \right\| T^{-1} \sum_{t=1}^T \|\vartheta_t\| \|l(D_t, z_t, \psi_0)\| \\ &\quad + \left\| \hat{p}(\hat{\psi}) \hat{\Omega}_{\hat{\psi}}^{-1} - \dot{p}(\psi_0) \Omega_{\hat{\psi}}^{-1} \right\| T^{-1} \sum_{t=1}^T \|\vartheta_t\| |B_t| \left\| \hat{\psi} - \psi_0 \right\|^\alpha \end{aligned}$$

where  $E \left[ (\|\vartheta_t\| \|l(D_t, z_t, \psi_0)\|)^{(p+\epsilon)/2} \right] < \infty$  as before. Then,  $T^{-1} \sum_{t=1}^T \|\vartheta_t\| \|l(D_t, z_t, \psi_0)\| = O_p(1)$  and (33), (34) and (42) imply that  $R_2 = o_p(1)$ .

For  $R_3$  note that

$$\begin{aligned} &\left\| T^{-1} \sum_{t=1}^T \vartheta_t \left( \dot{p}(\psi_0) \Omega_{\hat{\psi}}^{-1} \left( l(D_t, z_t, \hat{\psi}) - l(D_t, z_t, \psi_0) \right) \right)' \right\| \\ &\leq \left\| \dot{p}(\psi_0) \Omega_{\hat{\psi}}^{-1} \right\| T^{-1} \sum_{t=1}^T \|\vartheta_t\| \left\| l(D_t, z_t, \hat{\psi}) - l(D_t, z_t, \psi_0) \right\| \\ &\quad \left\| \dot{p}(\psi_0) \Omega_{\hat{\psi}}^{-1} \right\| T^{-1} \sum_{t=1}^T \|\vartheta_t\| |B_t| \left\| \hat{\psi} - \psi_0 \right\| \end{aligned}$$

where  $\left\| \hat{\psi} - \psi_0 \right\| = o_p(1)$  by Condition 4. Then,  $R_3 = o_p(1)$  follows from (42). The term  $T^{-1} \sum_{t=1}^T \left( \hat{\vartheta}_t - \vartheta_t \right) \times \left( \hat{\vartheta}_t - \vartheta_t \right)'$  in (36) can be analyzed in the same way as  $T^{-1} \sum_{t=1}^T \vartheta_t \left( \hat{\vartheta}_t - \vartheta_t \right)'$  but the details are omitted. It follows that  $T^{-1} \sum_{t=1}^T \left( \hat{\vartheta}_t \hat{\vartheta}_t' - \vartheta_t \vartheta_t' \right) = o_p(1)$  which in turn implies that

$$\hat{V} - V = o_p(1). \quad (43)$$

Then, for  $\bar{m} = T^{-1} \sum_{t=1}^T \hat{m}_t$ , the statistic  $T \bar{m}' \hat{V}^{-1} \bar{m}$  is asymptotically  $\chi_J^2$  because of (32), (43) and the continuous mapping theorem. ■

## References

- ANGRIST, J. D., AND G. W. IMBENS (1995): “Two-Stage Least Squares Estimates of Average Causal Response in Models with Variable Treatment Intensity,” *Journal of the American Statistical Association*, 90(430), 431–442.
- ANGRIST, J. D., AND G. M. KUERSTEINER (2004): “Semiparametric Causality Tests Using the Policy Propensity Score,” *NBER Working Paper 10975*.
- ANGRIST, J. D., AND G. M. KUERSTEINER (2011): “Causal effects of monetary shocks: Semiparametric conditional independence tests with a multinomial propensity score,” *Review of Economics and Statistics*.
- BAGLIANO, F. C., AND C. A. FAVERO (1998): “Measuring Monetary Policy with VAR Models: An Evaluation,” *European Economic Review*, 42(6), 1069–1112.
- BERNANKE, B. S., AND K. N. KUTTNER (2005): “What Explains the Stock Market’s Reaction to Federal Reserve Policy?,” *Journal of Finance*, 60(3), 1221–1257.
- BLANCHARD, O., AND J. GALÍ (2010): “Labor Markets and Monetary Policy: A New Keynesian Model with Unemployment,” *American Economic Journal: Macroeconomics*, 2, 1–30.
- CHRISTIANO, L. J., M. EICHENBAUM, AND C. EVANS (1996): “The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds,” *The Review of Economics and Statistics*, 78, 16–34.
- CHRISTIANO, L. J., M. EICHENBAUM, AND C. L. EVANS (1999): “Monetary Policy Shocks: What Have We Learned And To What End?,” in *Handbook of Macroeconomics, Volume 1*, ed. by J. B. Taylor, and M. Woodford, chap. 2, pp. 65–148. Elsevier Science B.V.
- CHRISTIANO, L. J., M. TRABANDT, AND K. WALENTIN (2011): “DSGE Models for Monetary Policy Analysis,” in *Handbook of Monetary Economics, Volume 3A*, ed. by B. M. Friedman, and M. Woodford, pp. 285–367. North Holland, Netherlands.
- CLARIDA, R., J. GALÍ, AND M. GERTLER (2000): “Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory,” *The Quarterly Journal of Economics*, pp. 147–180.
- COCHRANE, J. H. (1994): “Comment on: What Ends Recessions? by Christina D. Romer and David H. Romer,” in *NBER Macroeconomics Annual*, ed. by S. Fischer, and J. J. Rotemberg, vol. 9. MIT Press: Cambridge.
- COCHRANE, J. H., AND M. PIAZZESI (2002): “The Fed and Interest Rates - A High Frequency Identification,” *The American Economic Review*, 92, 90–95.

- COVER, J. P. (1992): “Asymmetric Effects of Positive and Negative Money-Supply Shocks,” *The Quarterly Journal of Economics*, 107(4), 1261–1282.
- DEHEJIA, R. H., AND S. WAHBA (1999): “Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs,” *Journal of the American Statistical Association*, 94, 1053–1062.
- DELONG, J. B., AND L. H. SUMMERS (1988): “How Does Macroeconomic Policy Affect Output?,” *Brookings Papers on Economic Activity*, 2, 433–94.
- FAUST, J., E. T. SWANSON, AND J. H. WRIGHT (2004): “Identifying VARS based on high frequency data,” *Journal of Monetary Economics*, 51, 1107–1131.
- FRIEDMAN, M., AND A. J. SCHWARTZ (1963): *A Monetary History of the United States: 1867-1960*. National Bureau of Economic Research. Princeton University Press: Princeton, NJ.
- GALÍ, J. (2011): “Monetary Policy and Unemployment,” in *Handbook of Monetary Economics, Volume 3A*, chap. 10, pp. 487–546. Elsevier Science B.V.
- GALÍ, J., AND M. GERTLER (2007): “Macroeconomic Modeling for Monetary Policy Evaluation,” *Journal of Economic Perspectives*, 21, 25–45.
- GALLANT, A. R., P. E. ROSSI, AND G. TAUCHEN (1993): “Nonlinear Dynamic Structures,” *Econometrica*, 61(4), 871–907.
- GRANGER, C. W. J. (1969): “Investigating Causal Relations by Econometric Models and Cross Spectral Methods,” *Econometrica*, 37(3), 424–438.
- GURKAYNAK, R. S., B. P. SACK, AND E. SWANSON (2005): “The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models,” *American Economic Review*, 95(1), 425–436.
- GURKAYNAK, R. S., B. P. SACK, AND E. T. SWANSON (2007): “Market-Based Measures of Monetary Policy Expectations,” *Journal of Business and Economic Statistics*, 25(2), 201–212.
- HALL, P., AND C. HEYDE (1980): *Martingale Limit Theory and its Application*. Academic Press.
- HAMILTON, J. D. (2009): “Daily Changes in Fed Funds Futures Prices,” *Journal of Money, Credit and Banking*, 41(4), 567–582.
- HAMILTON, J. D., AND ÒSCAR JORDÀ (2002): “A Model of the Federal Funds Rate Target,” *Journal of Political Economy*, 110, 1135–1167.
- HARRISON, H. M., AND D. M. KREPS (1979): “Martingales and Arbitrage in Multiperiod Securities Markets,” *Journal of Economic Theory*, 20, 381–408.

- HECKMAN, J. J., H. ICHIMURA, AND P. E. TODD (1998): “Matching as an Econometric Evaluation Estimator,” *Review of Economic Studies*, 65, 261–294.
- HIRANO, K., G. IMBENS, AND G. RIDDER (2003): “Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score,” *Econometrica*, 71, 1161–1189.
- HOOVER, K. D., AND S. J. PEREZ (1994): “Post hoc ergo propter hoc once more: An evaluation of ‘Does monetary policy matter?’ in the spirit of James Tobin,” *Journal of Monetary Economics*, 34, 47–74.
- HORVITZ, D., AND D. THOMPSON (1952): “A Generalization of Sampling Without Replacement from a finite Population,” *Journal of the American Statistical Association*, 47(260), 663–685.
- IMBENS, G. W. (2000): “The Role of the Propensity Score in Estimating Dose-Response Functions,” *Biometrika*, 87, 706–710.
- JORDÀ, O. (2005): “Estimation and Inference of Impulse Responses by Local Projections,” *American Economic Review*, 95(1), 161–182.
- KAREKEN, J., AND R. M. SOLOW (1963): “Lags in Monetary Policy,” in *Commission on money and credit, stabilization policies.*, pp. 14–96. Englewood Cliffs, NJ: Prentice-Hall.
- KARRAS, G. (1996): “Are the Output Effects of Monetary Policy Asymmetric?,” *Oxford Bulletin of Economics and Statistics*, 58(2), 267–78.
- KEYNES, J. M. (1936): *The General Theory of Employment, Interest and Money*. McMillan.
- KRUEGER, J. T., AND K. N. KUTTNER (1996): “The Fed Funds Futures Rate as a Predictor of Federal Reserve Policy,” *The Journal of Futures Markets*, 16(8), 865–879.
- KUTTNER, K. N. (2001): “Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market,” *Journal of Monetary Economics*, 47(3), 523–544.
- LEEPER, E. M., AND T. ZHA (2003): “Modest Policy Intervention,” *Journal of Monetary Economics*, 50, 1673–1700.
- MAGNUS, J. R., AND H. NEUDECKER (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley and Sons.
- MCLEISH, D. (1975a): “Invariance Principles for Dependent Variables,” *Zeitschrift f. Wahrscheinlichkeitsrechnung und verw. Gebiete*, 32, 165–178.
- MCLEISH, D. L. (1975b): “A Maximal Inequality and Dependent Strong Laws,” *The Annals of Probability*, 3(5), 829–839.

- NEWKEY, W. K., AND K. D. WEST (1994): “Automatic Lag Selection in Covariance Matrix Estimation,” *Review of Economic Studies*, 61, 631–654.
- PIAZZESI, M. (2005): “Bond Yields and the Federal Reserve,” *Journal of Political Economy*, 113(2), 311–344.
- PIAZZESI, M., AND E. T. SWANSON (2008): “Futures Prices as Risk-Adjusted Forecasts of Monetary Policy,” *Journal of Monetary Economics*, 55(4), 677–691.
- RAVN, M. O., AND M. SOLA (2004): “Asymmetric Effects of Monetary Policy in the United States,” *Federal Reserve Bank of St. Louis Review*, 86(5), 41–60.
- ROMER, C. D., AND D. H. ROMER (1989): “Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz,” *NBER Macroeconomics Annual*, 4, 121–170.
- (1994): “Monetary Policy Matters,” *Journal of Monetary Economics*, 34(1), 75–88.
- ROSENBAUM, P. R., AND D. B. RUBIN (1983): “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika*, 70(1), 41–55.
- SCOTTI, C. (2011): “A Bivariate Model of Federal Reserve and ECB Main Policy Rates,” *International Journal of Central Banking*, 7(3), 37–78.
- SIMS, C. A. (1972): “Money, Income and Causality,” *American Economic Review*, 62(4), 540–562.
- (1992): “Interpreting the macroeconomic time series facts: The effects of monetary policy,” *European Economic Review*, 36(5), 975–1000.
- SINGLETON, K. J. (2001): “Estimation of affine asset pricing models using the empirical characteristic function,” *Journal of Econometrics*, 102(1), 111–141.
- S ODERLIND, P., AND L. SVENSSON (1997): “New techniques to extract market expectations from financial instruments,” *Journal of Monetary Economics*, 40, 383–429.
- TENREYRO, S., AND G. THWAITES (2013): “Pushing on a string: US monetary policy is less powerful in recessions,” working paper.
- THAPAR, A. (2008): “Using Private Information to Estimate the Effects of Monetary Policy,” *Journal of Monetary Economics*, 55, 806–824.
- WINGENDER, A. M. (2011): “Monetary Policy Shocks and Risk Premia in the Interbank Market,” *The B.E. Journal of Macroeconomics*, 11(1), 1–19.
- WOODFORD, M. (2001): “The Taylor Rule and Optimal Monetary Policy,” *American Economic Review*, 91(2), 232–237.

	July 1989 to July 2005				July 1989 to Dec 2008			
	$OP_{T1}$ (1)	$OP_{T2}$ (2)	$OP_{F1}$ (3)	$OP_{F2}$ (4)	$OP_{T1}$ (5)	$OP_{T2}$ (6)	$OP_{F1}$ (7)	$OP_{F2}$ (8)
<i>Meeting Months (<math>s_t^1</math>)</i>								
Pre-Crisis			1.08*** (0.25)	1.05*** (0.23)			1.18*** (0.23)	1.16*** (0.24)
Post-Crisis							0.86*** (0.28)	0.90*** (0.28)
<i>Non-Meeting Months (<math>s_t^0</math>)</i>								
Pre-Crisis			0.39 (0.28)	0.38 (0.29)			0.42 (0.30)	0.39 (0.30)
Post-Crisis							-0.00 (0.44)	-0.10 (0.47)
<i>Meeting months with pre-meeting changes (<math>s_t^0</math>)</i>								
Pre-Crisis			-0.11 (0.22)	-0.10 (0.23)			-0.12 (0.23)	-0.10 (0.24)
Inflation, Lag 1	0.00 (0.06)	0.02 (0.06)		0.02 (0.04)	0.02 (0.04)	0.01 (0.04)		0.01 (0.03)
Inflation, Lag 2		0.09 (0.06)		0.04 (0.05)		0.12** (0.05)		0.02 (0.03)
Unem. Rate, Lag 1	-0.31*** (0.10)	-0.19** (0.09)		-0.08 (0.07)	-0.30*** (0.09)	-0.17** (0.09)		-0.08 (0.06)
Unem. Rate, Lag 2		-0.25*** (0.09)		-0.05 (0.06)		-0.29*** (0.09)		-0.04 (0.06)
Target Rate		-0.01*** (0.01)	-0.01*** (0.00)	-0.02*** (0.00)		-0.02*** (0.01)	-0.01*** (0.00)	-0.02*** (0.00)
Last Target Change		0.11 (0.08)	0.13** (0.06)	0.11* (0.06)		0.11 (0.09)	0.15** (0.06)	0.13** (0.06)
LTC×FOMC		0.21** (0.10)	0.06 (0.07)	0.08 (0.07)		0.29*** (0.11)	0.09 (0.07)	0.11 (0.07)
FOMC		-0.02 (0.03)	-0.05* (0.03)	-0.04* (0.03)		-0.03 (0.03)	-0.05* (0.02)	-0.04* (0.02)
CRISIS					-0.07* (0.04)	-0.06* (0.03)	0.03 (0.03)	0.03 (0.03)
Log Likelihood	-187.65	-150.11	-102.92	-101.38	-232.55	-185.80	-119.25	-118.18
Pseudo $R^2$	0.01	0.15	0.37	0.38	0.00	0.15	0.43	0.43
Observations			192				233	

Table 1. Ordered Probit Specifications for the expected change of the Target Rate. This table reports selected marginal effects on the probability of a 25 bps increase in the fed funds target rate. Standard errors are shown in parentheses. \*\*\*/\*\*/\* indicates significance at the 99/95/90% confidence level. LTC=Last Target Change. For other variable definitions see text.

	July 1989 to July 2005				July 1989 to Dec 2008			
	$OP_{T1}$ (1)	$OP_{T2}$ (2)	$OP_{F1}$ (3)	$OP_{F2}$ (4)	$OP_{T1}$ (5)	$OP_{T2}$ (6)	$OP_{F1}$ (7)	$OP_{F2}$ (8)
A. Propensity Score Covariates								
FFF <sub>t</sub> Combined	0.000***	0.004***	0.434	0.465	0.000***	0.000***	0.299	0.288
FFF <sub>t</sub>	0.017**	0.061*	0.461	0.479	0.005***	0.108	0.195	0.188
FFF <sub>t</sub> squared	0.001***	0.001***	0.130	0.155	0.000***	0.000***	0.116	0.112
Effective Funds Rate	0.031**	0.114	0.256	0.262	0.019**	0.058*	0.127	0.130
Target Rate	0.014**	0.107	0.248	0.265	0.008***	0.058*	0.145	0.149
Last Target Change	0.000***	0.386	0.739	0.769	0.000***	0.087*	0.620	0.631
Inflation, Lag 1	0.913	0.952	0.966	0.998	0.461	0.430	0.753	0.764
Unem. Rate Lag 1	0.220	0.673	0.790	0.697	0.661	0.974	0.957	0.808
Inflation, Lag 2	0.896	0.490	0.714	0.732	0.881	0.694	0.712	0.771
Unem. Rate Lag 2	0.003***	0.720	0.252	0.394	0.007***	0.930	0.647	0.845
B. Lagged Outcome Variables								
Federal Funds Rate	0.000***	0.133	0.545	0.631	0.000***	0.031**	0.602	0.648
IP	0.193	0.439	0.848	0.896	0.175	0.398	0.817	0.819
3 Month T-Bill	0.255	0.844	0.876	0.907	0.048**	0.456	0.646	0.705
2 Year T-Bond	0.168	0.871	0.977	0.940	0.142	0.776	0.982	0.958
10 Year T-Bond	0.423	0.985	0.890	0.864	0.674	0.908	0.815	0.773

Table 2. Specification Tests. This table shows p-values for joint tests of the null hypothesis that the ordered probit model correctly specifies the conditional probability of  $D_t = \{-.25, 0, .25\}$ . Moment conditions are formed for each candidate predictor variable individually. Test statistics labeled FFF (Combined) are joint for  $FFF$  and  $FFF^2$ .

	$P_T$	$P_{F1}$	$P_{F2}$
	(1)	(2)	(3)
$FF_t$		1.39***	1.93***
		(0.41)	(0.60)
Inflation	0.07		-0.37*
	(0.16)		(0.20)
Unemployment Rate	0.25		-0.42
	(0.58)		(0.54)
Log Likelihood	-15.99	-10.71	-8.73
Pseudo R2	0.06	0.45	0.58
Sample Size	28	28	28

Table 3. Probit Models for Target Rate Reductions during the 2006-2008 period. The table shows marginal effects. \*\*\*/\*\*/\* indicates significance at the 99/95/90% confidence level

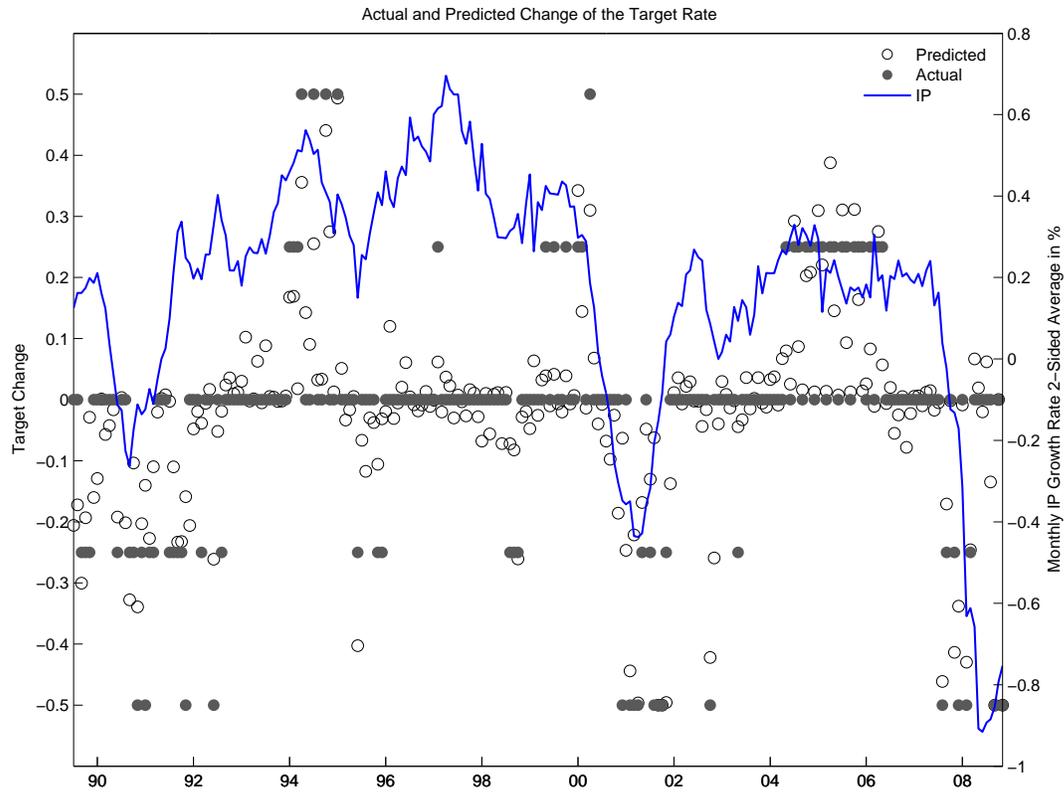


Figure 1. Actual Changes (circles) and Predicted Changes (dots) in the Target Rate. Predictions are from the propensity score model labeled  $OP_{F2}$  in Table 1. The figure also shows IP growth over the same period.

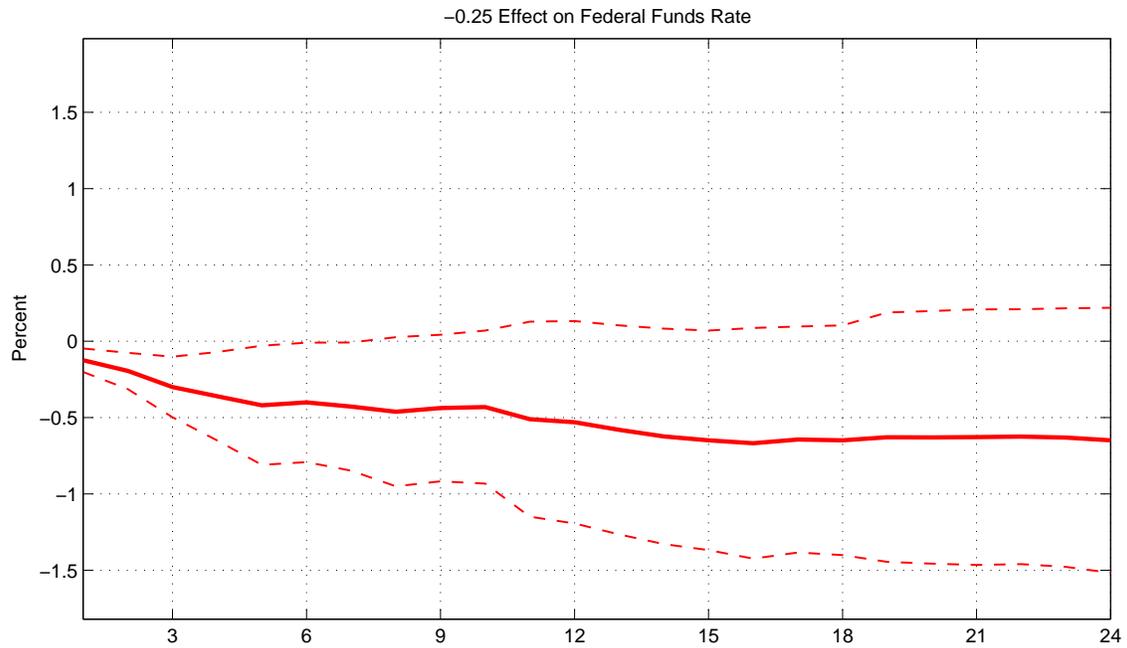
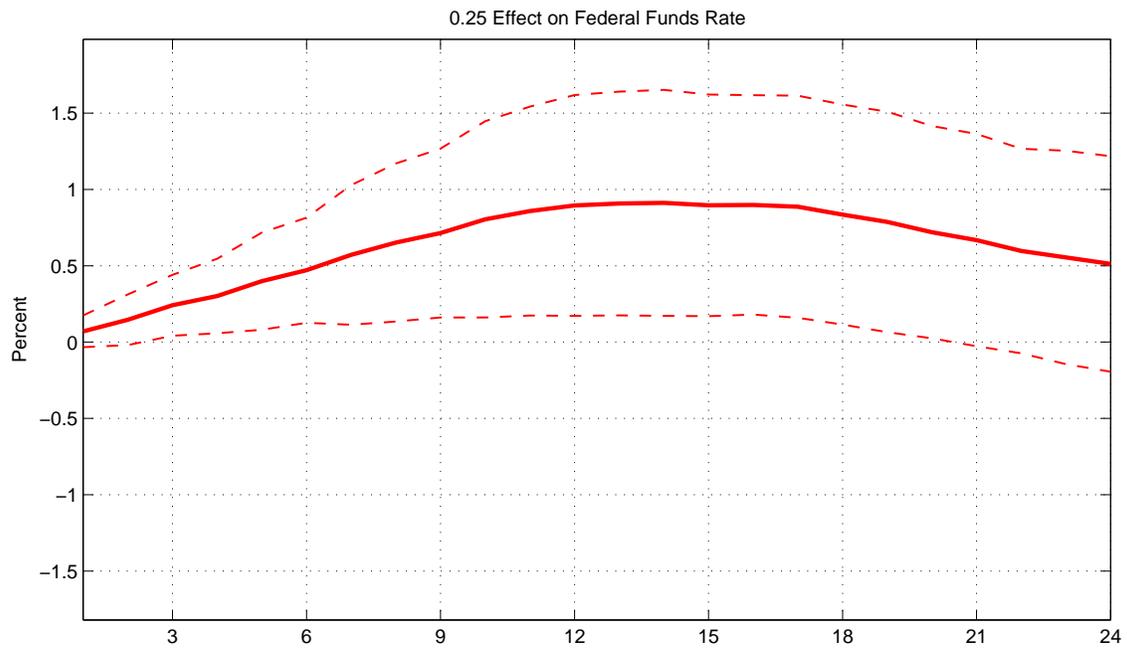


Figure 2. Estimated Effects of Target Rate Changes on the Federal Funds Rate. These estimates use data from August 1989 through July 2007, and the propensity score mode labeled  $OP_{F2}$  in Table 1. Dashed lines indicate 90% confidence bands.

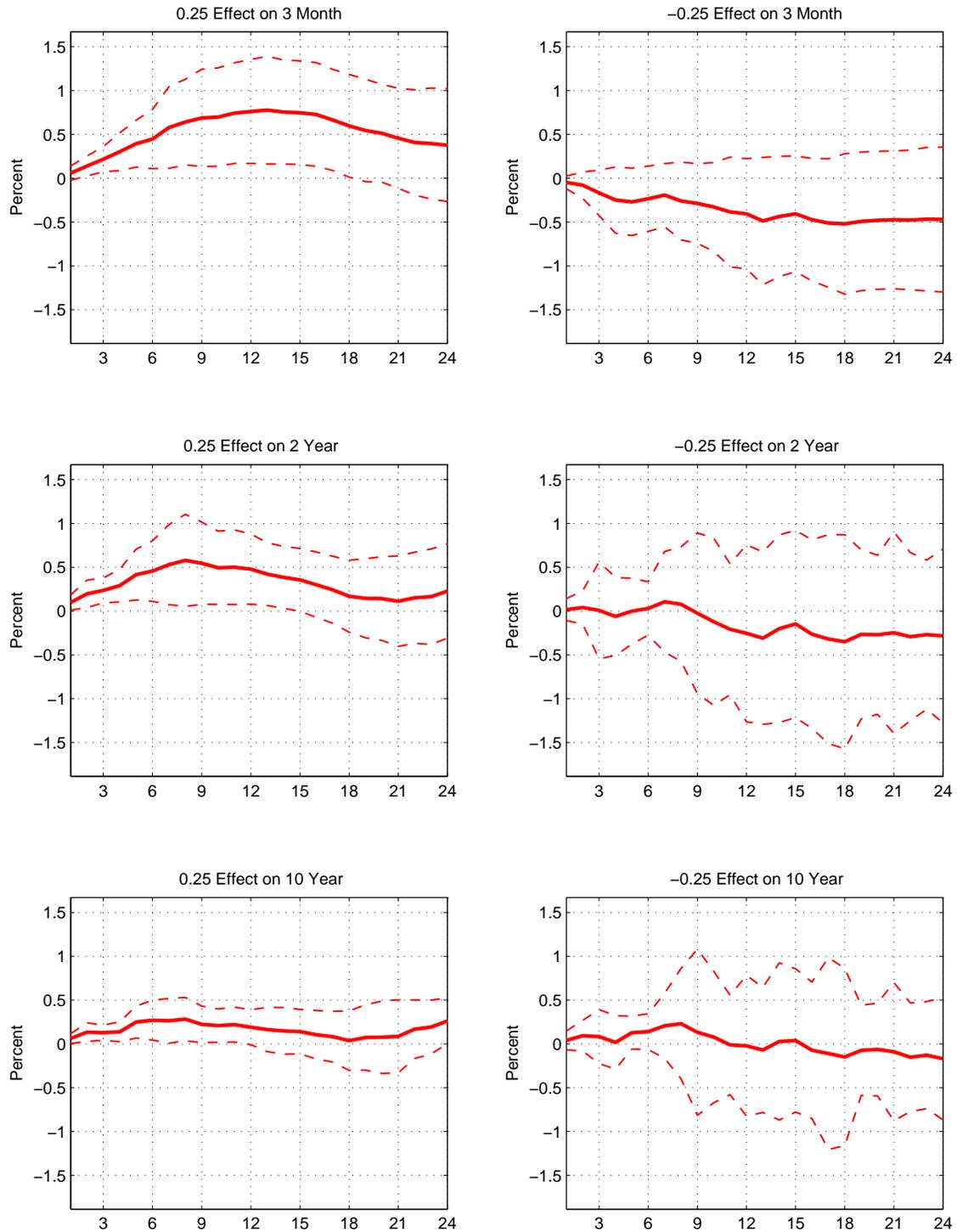


Figure 3. Estimated Effects of Target Rate Changes on Bond Yields. These estimates use data from August 1989 through July 2007, and the propensity score mode labeled  $OP_{F2}$  in Table 1. Dashed lines indicate 90% confidence bands.

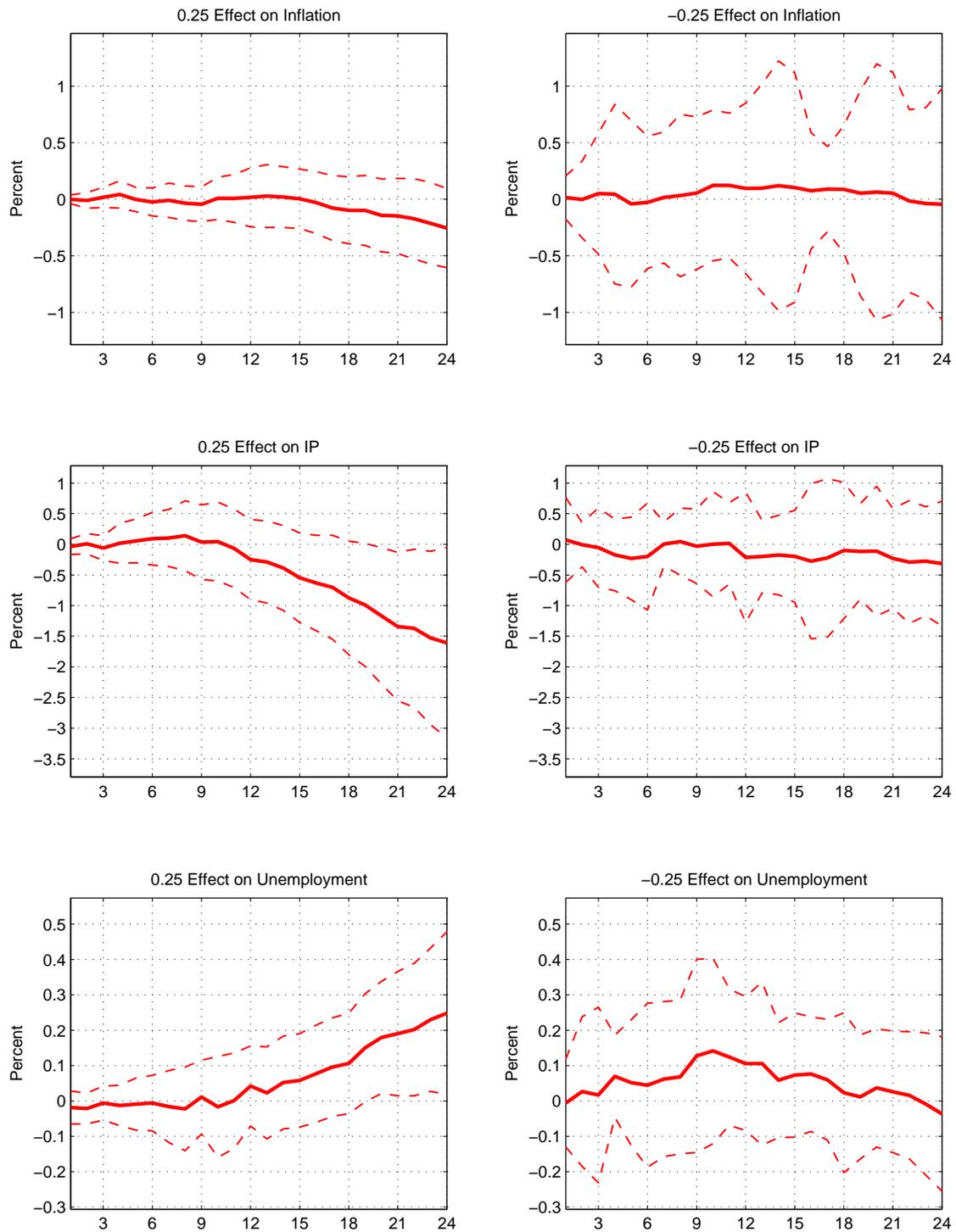


Figure 4. Estimated Effects of Target Rate Changes on Macro Variables. These estimates use data from August 1989 through July 2007, and the propensity score mode labeled  $OP_{F2}$  in Table 1. Dashed lines indicate 90% confidence bands.

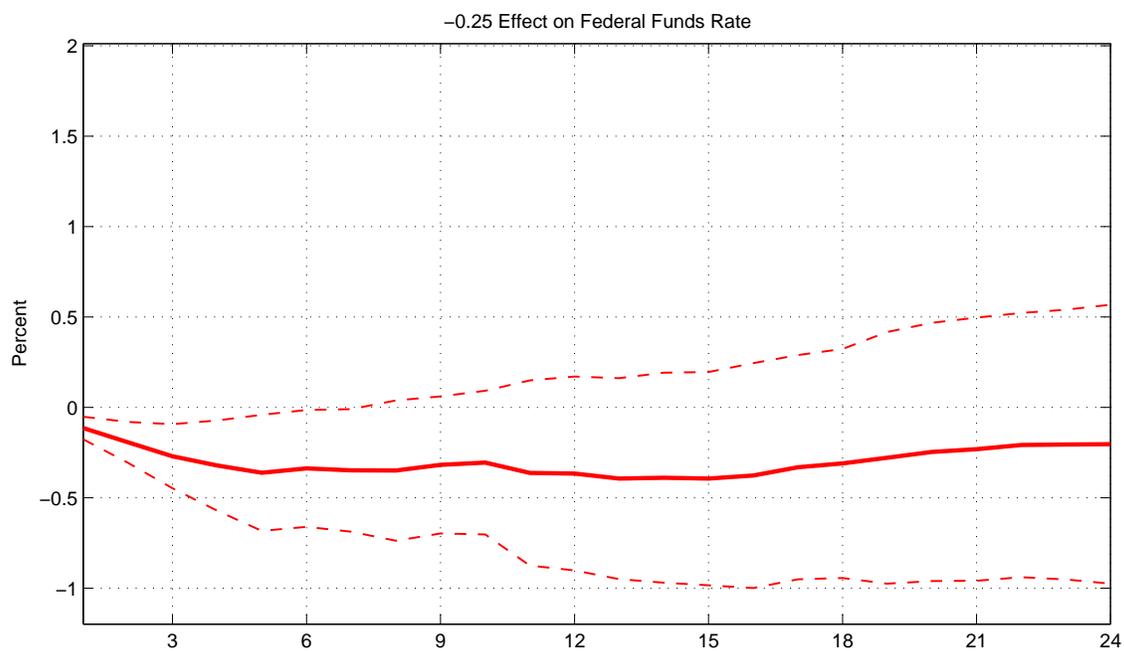
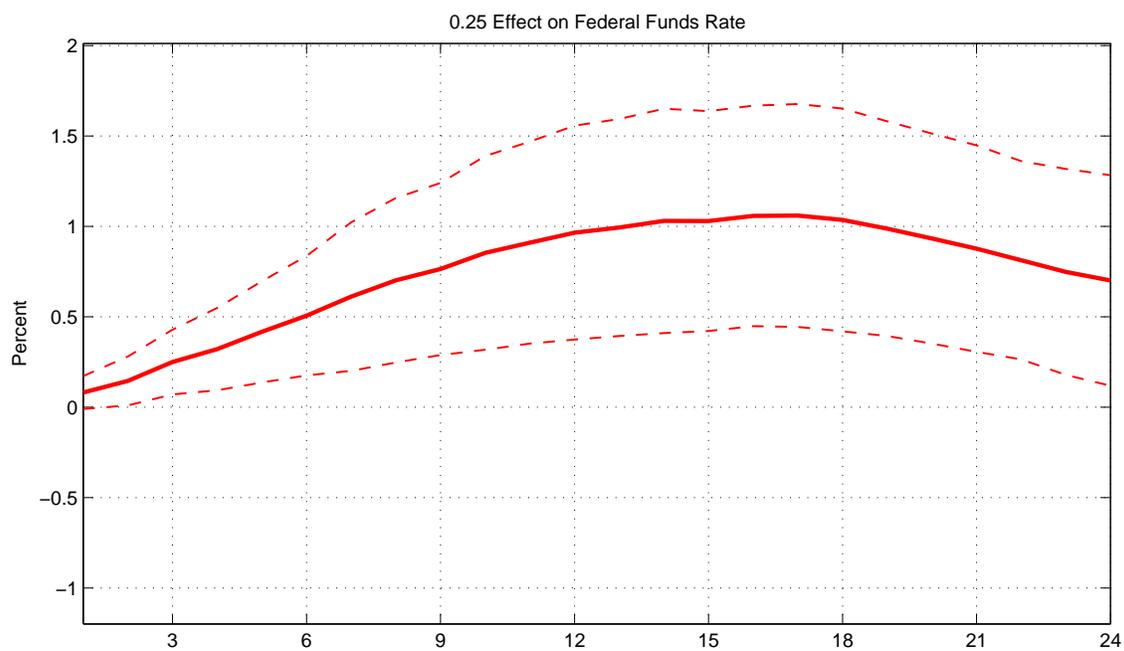


Figure 5. Estimated Effects of Target Rate Changes on the Federal Funds Rate Through 2010. These estimates use data from August 1989 through December 2010, and the propensity score model (8) labeled  $OP_{F2}$  in Table 1. Dashed lines indicate 90% confidence bands.

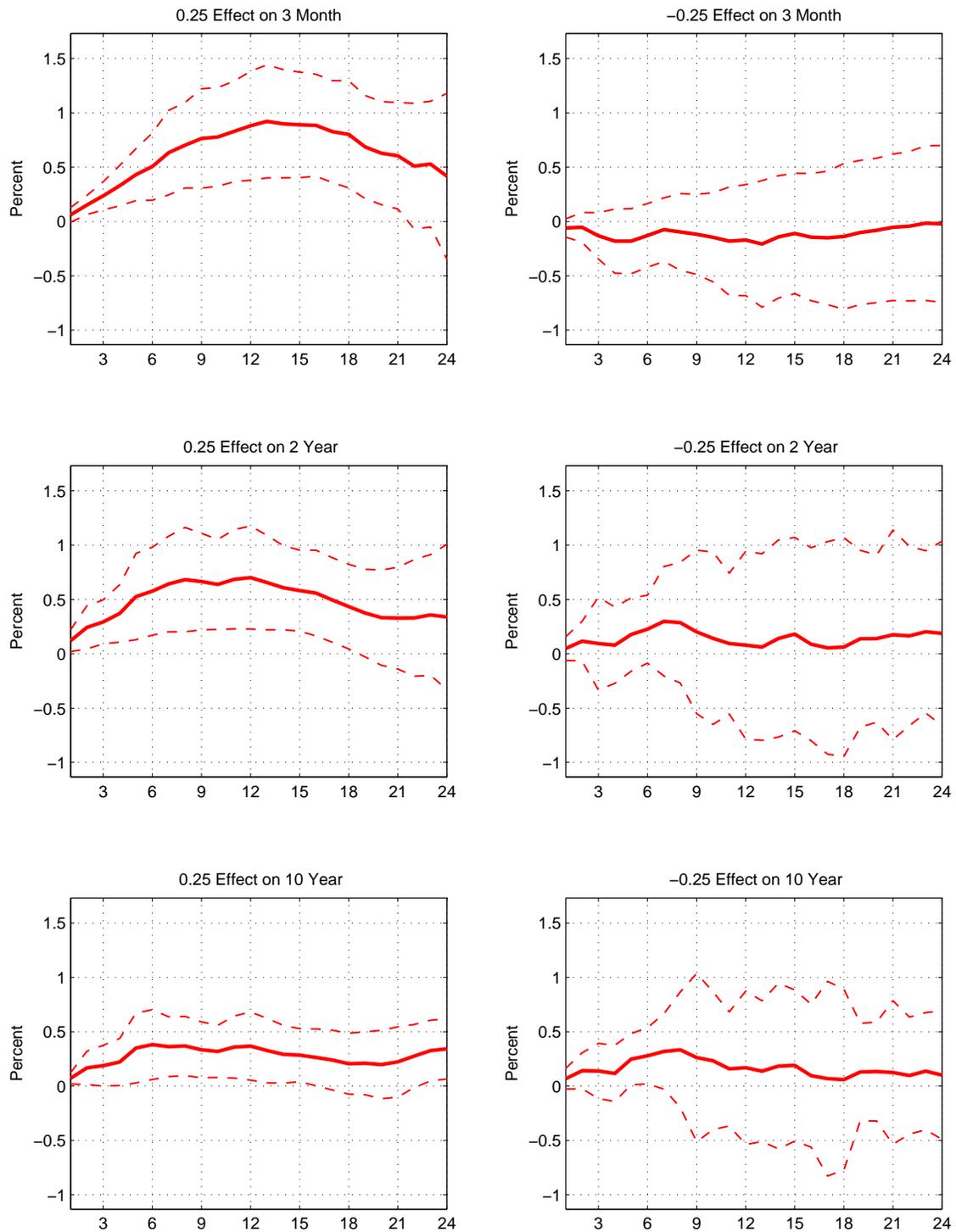


Figure 6. Estimated Effects of Target Rate Changes on Bond Yields Through 2010. These estimates use data from August 1989 through December 2010, and the propensity score model (8) labeled  $OP_{F2}$  in Table 1. Dashed lines indicate 90% confidence bands.

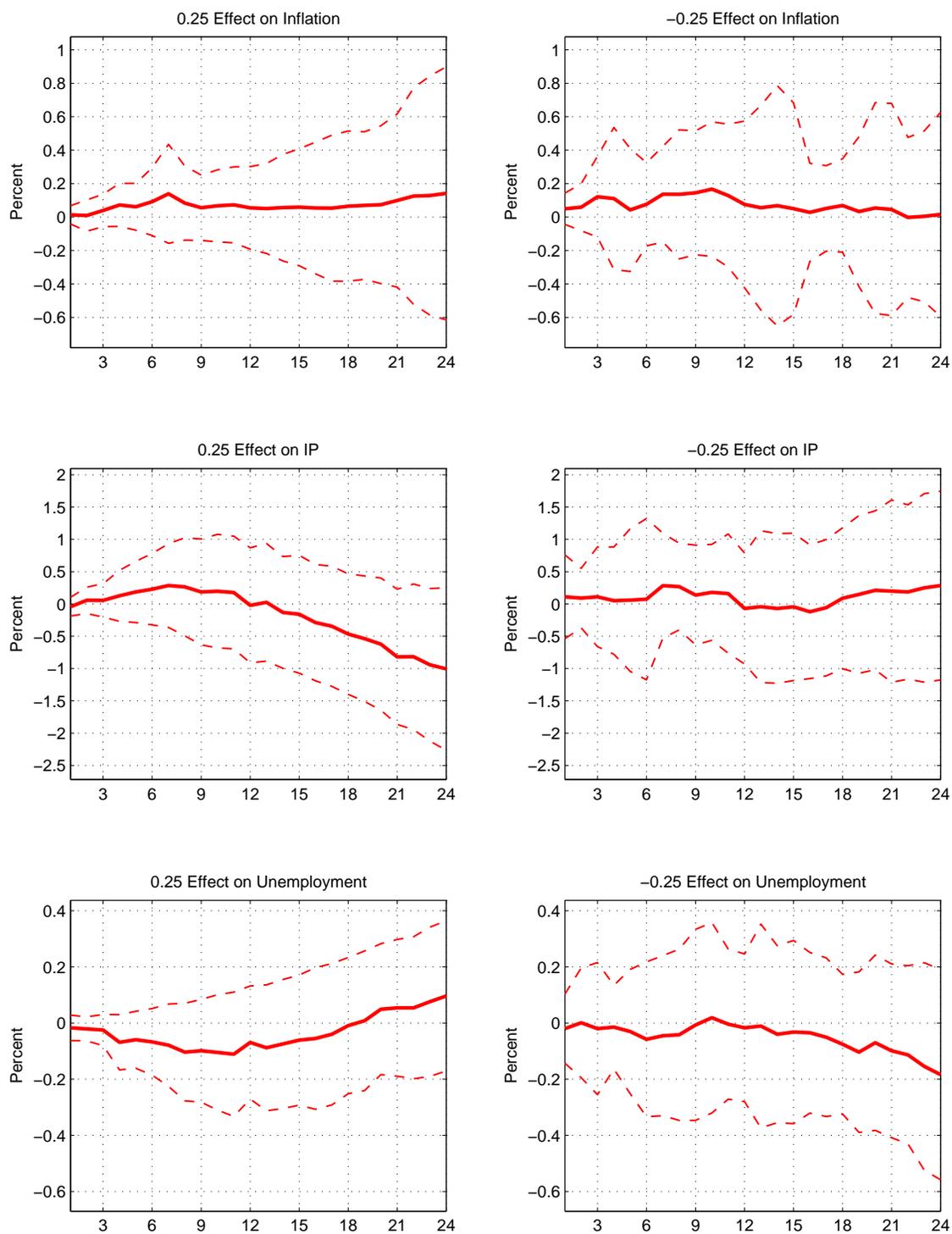


Figure 7. Estimated Effects of Target Rate Changes on Macro Variables Through 2010. These estimates use data from August 1989 through December 2010, and the propensity score model (8) labeled  $OP_{F2}$  in Table 1. Dashed lines indicate 90% confidence bands.

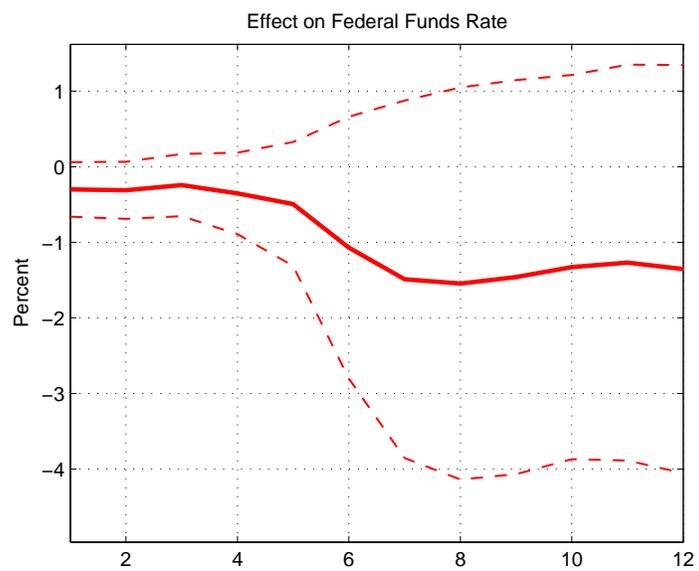


Figure 8. Estimated Effects of Target Rate Drops on the Federal Funds Rate. These estimates use data from October 2006 through December 2009, and the propensity score model labeled PF2 in Table 3. Dashed lines indicate 90% confidence bands.

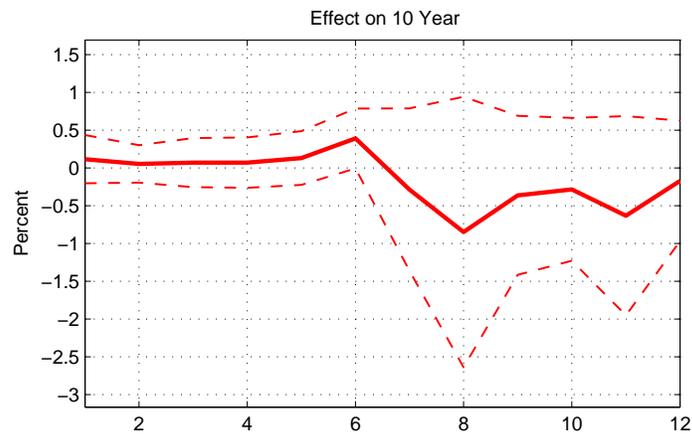
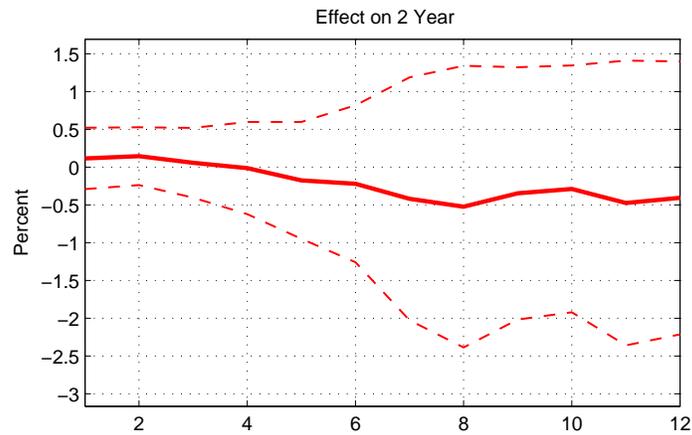
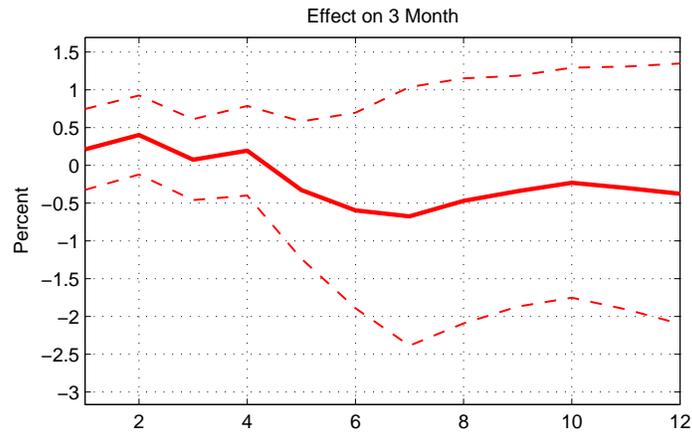


Figure 9. Estimated Effects of Target Rate Drops on Bond Yields in the Crisis Period. These estimates use data from October 2006 through December 2009, and the propensity score model labeled PF2 in Table 3. Dashed lines indicate 90% confidence bands.

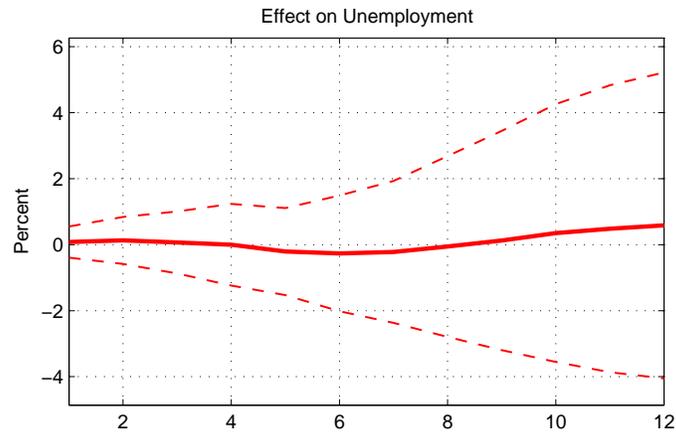
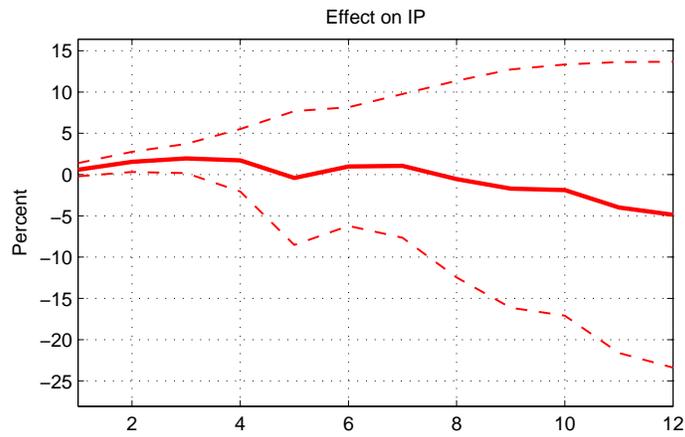
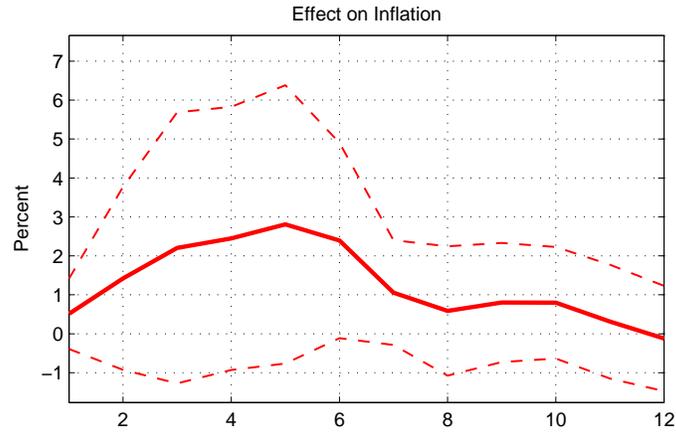


Figure 10. Estimated Effects of Target Rate Drops on Macro Variables in the Crisis Period. These estimates use data from October 2006 through December 2009, and the propensity score model labeled  $P_{F2}$  in Table 3. Dashed lines indicate 90% confidence bands.