

## MIT Open Access Articles

*Causal Effects of Monetary Shocks: Semiparametric Conditional Independence Tests with a Multinomial Propensity Score*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Angrist, Joshua D., and Guido M. Kuersteiner. "Causal effects of monetary shocks: Semiparametric conditional independence tests with a multinomial propensity score." Forthcoming in *Review of Economics and Statistics*

**As Published:** [http://dx.doi.org/10.1162/REST\\_a\\_00109](http://dx.doi.org/10.1162/REST_a_00109)

**Publisher:** MIT Press

**Persistent URL:** <http://hdl.handle.net/1721.1/61874>

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

**Terms of use:** Creative Commons Attribution-Noncommercial-Share Alike 3.0



# Causal effects of monetary shocks: Semiparametric conditional independence tests with a multinomial propensity score<sup>1</sup>

Joshua D. Angrist

Guido M. Kuersteiner

MIT

Georgetown University

This Revision: April 2010

<sup>1</sup>This paper is a revised version of NBER working paper 10975, December 2004. We thank the Romers' for sharing their data, Alberto Abadie, Xiaohong Chen, Jordi Gali, Simon Gilchrist, Yanqin Fan, Stefan Hoderlein, Bernard Salanié and James Stock for helpful discussions and seminar and conference participants at ASSA 2006, Auckland, Berkeley, BU, Columbia, Harvard-MIT, Georgetown, Indiana, Illinois, Konstanz, LAMES 2006, the 2004 NSF/NBER Time Series Conference, CIREQ Montreal, Rochester, Rutgers, Stanford, St. Gallen, the Triangle Econometrics Workshop, UCLA, Wisconsin, Zurich and the 2<sup>nd</sup> conference on evaluation research in Mannheim for helpful comments. Thanks also go to the editor and referees for helpful suggestions and comments. Kuersteiner gratefully acknowledges support from NSF grants SES-0095132 and SES-0523186.

## Abstract

We develop semiparametric tests for conditional independence in time series models of causal effects. Our approach is motivated by empirical studies of monetary policy effects. Our approach is semiparametric in the sense that we model the process determining the distribution of treatment – the policy propensity score – but leave the model for outcomes unspecified. A conceptual innovation is that we adapt the cross-sectional potential outcomes framework to a time series setting. We also develop root-T consistent distribution-free inference methods for full conditional independence testing, appropriate for dependent data and allowing for first-step estimation of the (multinomial) propensity score.

**Keywords:** Potential outcomes, monetary policy, causality, conditional independence, functional martingale difference sequences, Khmaladze transform, empirical Rosenblatt transform

# 1 Introduction

The causal connection between monetary policy and real economic variables is one of the most important and widely studied questions in macroeconomics. Most of the evidence on this question comes from regression-based statistical tests. That is, researchers regress an outcome variable such as industrial production on measures of monetary policy, while controlling for lagged outcomes and contemporaneous and lagged covariates, with the statistical significance of policy variables providing the test results of interest. Two of the most influential empirical studies in this spirit are by Sims (1972, 1980), who discusses conceptual as well as empirical problems in the money-income nexus.

The foundation of regression-based causality tests is a simple conditional independence assumption. The core null hypothesis is that conditional on lagged outcomes and an appropriate set of control variables, the absence of a causal relationship should be manifest in a statistically insignificant connection between policy surprises and contemporaneous and future outcomes. In the language of cross-sectional program evaluation, policy variables are assumed to be “as good as randomly assigned” after appropriate regression conditioning, so that conditional effects have a causal interpretation. While this is obviously a strong assumption, it seems like a natural place to begin empirical work, at least in the absence of a randomized trial or compelling exclusion restrictions. The conditional independence assumption is equivalent to postulating independent structural innovations in a structural vector autoregression (SVAR), a tool that has taken center stage in the analysis of monetary policy effects. Recent contributions to this literature include Bernanke and Blinder (1992), Christiano, Eichenbaum and Evans (1996, 1999), Gordon and Leeper (1994), Sims and Zha (2006), and Strongin (1995).

While providing a flexible tool for the analysis of causal relationships, an important drawback of

regression-based conditional independence tests, including those based on SVAR's, is the need for an array of auxiliary assumptions that are hard to assess and interpret, especially in a time series context. Specifically, regression tests rely on a model of the process determining GDP growth or other macroeconomic outcomes. Much of the recent literature in monetary macroeconomics has focused on dynamic stochastic general equilibrium (DSGE) models for this purpose. As discussed by Sims and Zha (2006), SVAR's can be understood as first-order approximations to a potentially non-linear DSGE model. Moreover, as a framework for hypothesis testing, the SVAR approach implicitly requires specification of both a null and an alternative model.

The principal contribution of this paper is to develop an approach to time series causality testing that shifts the focus away from a model of the process determining outcomes towards a model of the process determining policy decisions. In particular, we develop causality tests that rely on a model for the conditional probability of a policy shift, which we call the "policy propensity score", leaving the model for outcomes unspecified. In the language of the SVAR literature, our approach reduces the modeling burden to the specification, identification, and estimation of the structural policy innovation while leaving the rest of the system unspecified. This limited focus should increase robustness. For example, we do not need to specify functional form or lag length in a model for GDP growth. Rather, we need be concerned solely with the time horizon and variables relevant for Federal Open market Committee (FOMC) decision-making, issues about which there is some institutional knowledge. Moreover, the multinomial nature of policy variables such as the one we study provides a natural guide as to the choice of functional form for the policy model.

A second contribution of our paper is the outline of a potential-outcomes framework for causal research

using time series data. In particular, we show that a generalized Sims-type definition of dynamic causality provides a coherent conceptual basis for time series causal inference analogous to the selection-on-observables assumption widely used in cross-section econometrics. The analogy between a time series causal inquiry and a cross-sectional selection-on-observables framework is even stronger when the policy variable can be coded as a discrete treatment-type variable. In this paper, therefore, we focus on the causal effect of changes in the federal funds target rate, which tends to move up or down in quarter-point jumps. Our empirical work is motivated by Romer and Romer's (2004) analysis of the FOMC decisions regarding the intended federal funds rate. This example is also used to make our theoretical framework concrete. In an earlier paper, Romer and Romer (1989) described monetary policy shocks using a dummy variable for monetary tightening. An application of our framework to this binary-treatment case appears in our working paper (Angrist and Kuersteiner, 2004). Here, we consider a more general model of the policy process where Federal Funds target rate changes are modeled as a dynamic multinomial process.

Propensity score methods, introduced by Rosenbaum and Rubin (1983), are now widely used for cross-sectional causal inference in applied econometrics. Important empirical examples include Dehejia and Wahba (1999) and Heckman, Ichimura and Todd(1998), both of which are concerned with evaluation of training programs. Heckman, Ichimura, and Todd (1997), Heckman, et al (1998), and Abadie (2005) develop propensity score strategies for differences-in-differences estimators. The differences-in-differences framework often has a dynamic element since these models typically involve intertemporal comparisons. Similarly, Robins, Greenland and Hu (1999), Lok et.al. (2004) and Lechner (2004) have considered panel-type settings with time-varying treatments and sequential randomized trials. At the same time, few, if any, studies have considered propensity score methods for a pure time series application. This in spite

of the fact that the dimension-reducing properties of propensity score estimators would seem especially attractive in a time series context. Finally, we note that Imbens (2000) and Lechner (2000) generalize the binary propensity score approach to allow for ordered treatments, though this work has not yet featured widely in applications.

Implementation of our semiparametric test for conditional independence in time series data generates a number of inference problems. First, as in the cross-sectional and differences-in-differences settings discussed by Hahn (1999), Heckman, Ichimura and Todd (1998), Hirano, Imbens, and Ridder (2003), Abadie (2005), and Abadie and Imbens (2009), inference should allow for the fact that in practice the propensity score is unknown and must be estimated. First-step estimation of the propensity score changes the limiting distribution of our Kolmogorov-Smirnov (KS) and von Mises (VM) test statistics.

A second and somewhat more challenging complication arises from the fact that non-parametric tests of distributional hypotheses such as conditional independence may have a non-standard limiting distribution, even in a relatively simple cross-sectional setting. For example, in a paper closely related to ours, Linton and Gozalo (1999) consider KS- and VM-type statistics, as we do, but the limiting distributions of their test statistics are not asymptotically distribution-free, and must therefore be bootstrapped.<sup>1</sup> More recently, Su and White (2003) propose a nonparametric conditional independence test for time series data based on orthogonality conditions obtained from an empirical likelihood specification. The Su and White procedure converges at a less-than-standard rate due to the need for nonparametric density estimation. In contrast, we present new Kolmogorov-Smirnov (KS) and von Mises (VM) statistics that provide distribution-free tests for full conditional independence, are suitable for dependent data, and

---

<sup>1</sup>See also Abadie (2002), who proposes a bootstrap procedure for nonparametric testing of hypotheses about the distribution of potential outcomes, when the latter are estimated using instrumental variables.

which converge at the standard rate.

The key to our ability to improve on previous tests of conditional independence, and an added benefit of the propensity score, is that we are able to reduce the problem of testing for conditional distributional independence to a problem of testing for a martingale difference sequence (MDS) property of a certain functional of the data. This is related to the problem of testing for the MDS property of simple stochastic processes, which has been analyzed by, among others, Bierens (1982, 1990), Bierens and Ploberger (1997), Chen and Fan (1999), Stute, Thies and Zhu (1998) and Koul and Stute (1999). Our testing problem is more complicated because we simultaneously test for the MDS property of a continuum of processes indexed in a function space. Earlier contributions propose a variety of schemes to find critical values for the limiting distribution of the resulting test statistics but most of the existing procedures involve nuisance parameters.<sup>2</sup> Our work extends Koul and Stute (1999) by allowing for more general forms of dependence, including mixing and conditional heteroskedasticity. These extensions are important in our application because even under the null hypothesis of no causal relationship, the observed time series are not Markovian and do not have a martingale difference structure. Most importantly, direct application of the Khmaladze (1988,1993) method in a multivariate context appears to work poorly in practice. We therefore use a Rosenblatt (1952) transformation of the data in addition to the Khmaladze transformation<sup>3</sup>. This combination of methods seems to perform well, at least for the low-dimensional

---

<sup>2</sup>In light of this difficulty, Bierens and Ploberger (1997) propose asymptotic bounds, Chen and Fan (1999) use a bootstrap and Koul and Stute (1999) apply the Khmaladze transform to produce a statistic with a distribution-free limit. The univariate version of the Khmaladze transform was first used in econometrics by Bai (2003) and Koenker and Xiao (2002).

<sup>3</sup>In recent work, independent of ours, Delgado and Stute (2008) discuss a specification test that also combines the Khmaladze and Rosenblatt transforms. Song (2009) considers a nonparametric test (as opposed to our semiparametric



multivariate systems explored here.

The paper is organized as follows. The next section outlines our conceptual framework, while Section 3 provides a heuristic derivation of the testing strategy. Section 4 discusses the construction of feasible critical values using the Khmaladze and Rosenblatt transforms as well as a bootstrap procedure. Finally, the empirical behavior of alternative causality concepts and test statistics is illustrated through a re-analysis of the Romer and Romer (2004) data in Section 5.<sup>4</sup> As an alternative to the Romers' approach, and to illustrate the use of our framework for specification testing, we also explore a model for monetary policy based on a simple Taylor rule. Appendix A extends the tests of Section 3 to a general testing framework. Appendix B provides detailed descriptions on how to implement the test statistics. Appendix C summarizes theoretical results and technical assumptions.<sup>5</sup> Appendices D and E contain model and data definitions for the empirical work in Section 5.

## 2 Notation and Framework

Causal effects are defined here using the Rubin (1974) notion of potential outcomes. The potential outcomes concept originated in randomized trials, but is now widely used in observational studies. Our definition of causality relies on the distinction between potential outcomes that would be realized with and without a change in policy. In the case of a binary treatment, these are denoted by  $Y_{1t}$  and  $Y_{0t}$ . The test) of conditional independence using the Rosenblatt transform. In his setup, parameter estimation involving conditioning variables, unlike in our case for the propensity score, does not affect the limiting distribution of test statistics. This eliminates the need for the Khmaladze transform.

---

<sup>4</sup>A small Monte Carlo study can be found in our NBER working paper Angrist and Kuersteiner (2004).

<sup>5</sup>Proofs are available in an Auxiliary Appendix published online.

observed outcome in period  $t$  can then be written  $Y_t = Y_{1t}D_t + (1 - D_t)Y_{0t}$ , where  $D_t$  is treatment status. In the absence of serial correlation or covariates, the causal effect of a treatment or policy action is defined as  $Y_{1t} - Y_{0t}$ . Since only one or the other potential outcome can ever be observed, researchers typically focus on the average causal effect  $E(Y_{1t} - Y_{0t})$ , or the average effect in treated periods,  $E(Y_{1t} - Y_{0t} | D_t = 1)$ . When  $D_t$  takes on more than two values, there are multiple incremental average treatment effects, e.g., the effect of going up or down. This is spelled out further below.

Time series data are valuable in that, by definition, a time series sample consists of repeated observations on the subject of interest (typically a country or economy). At the same time, time series application pose special problems for causal inference. In a dynamic setting, the definition of causal effects is complicated by the fact that potential outcomes are determined not just by current policy actions but also by past actions, lagged outcomes, and covariates. To capture dynamics, we assume the economy can be described by the *observed* vector stochastic process  $\chi_t = (Y_t, X_t, D_t)$ , defined on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , where  $Y_t$  is a vector of outcome variables,  $D_t$  is a vector of policy variables, and  $X_t$  is a vector of other exogenous and (lagged) endogenous variables that are not part of the null hypothesis of no causal effect of  $D_t$ . Let  $\bar{X}_t = (X_t, \dots, X_{t-k}, \dots)$  denote the covariate path, with similar definitions for  $\bar{Y}_t$  and  $\bar{D}_t$ . Formally, the relevant information is assumed to be described by  $\mathcal{F}_t = \sigma(z_t)$  where  $z_t = \Pi_t(\bar{X}_t, \bar{Y}_t, \bar{D}_{t-1})$  is a sequence of finite dimensional functions  $\Pi_t : \bigotimes_{i=1}^{\dim(\chi_t)} \mathbb{R}^\infty \rightarrow \mathbb{R}^{k_2}$  of the entire observable history of the joint process. For the purposes of empirical work, the mapping  $\Pi_t$  and  $z_t$  are assumed to be known.

A key to identification in our framework is the distinction between systematic and random components in the process by which policy is determined. Specifically, decisions about policy are assumed to be

determined in part by a possibly time-varying but non-stochastic function of observed random variables, denoted  $D(z_t, t)$ . This function summarizes the role played by observable variables in the policy makers' decision-making process. In addition, policy makers are assumed to react to idiosyncratic information, represented by the scalar  $\varepsilon_t$ , that is not observed by researchers and therefore modeled as a stochastic shock. The policy  $D_t$  is determined by both observed and unobserved variables according to  $D_t = \psi(D(z_t, t), \varepsilon_t, t)$ , where  $\psi$  is a general mapping. Without loss of generality we can assume that  $\varepsilon_t$  has a uniform distribution on  $[0, 1]$ . This is because  $\psi(a, b, t)$  can always be defined as  $\tilde{\psi}(a, F^{-1}(b), t)$  where  $F$  is any parametric or non-parametric distribution function. We assume that  $\psi$  takes values in the set of functions  $\Psi_t$ . A common specification in the literature on monetary policy is a Taylor (1993) rule for the nominal interest rate. In this literature,  $\psi$  is usually linear while  $z_t$  is lagged inflation and unemployment (see, e.g., Rotemberg and Woodford (1997)). A linear rule implicitly determines the distribution of  $\varepsilon_t$ .

A second key assumption is that the stochastic component of the policy function,  $\varepsilon_t$ , is independent of potential outcomes. This assumption is distinct from the policy model itself and therefore discussed separately, below. Given this setup, we can define potential outcomes as the possibly counterfactual realizations of  $Y_t$  that would arise in response to a hypothetical change in policy as described by alternative realizations for  $\psi(D(z_t, t), \varepsilon_t, t)$ . The definition allows counterfactual outcomes to vary with changes in policy realizations for a given policy rule, or for a changing policy rule:

**Definition 1** *A potential outcome,  $Y_{t,j}^\psi(d)$ , is defined as the value assumed by  $Y_{t+j}$  if  $D_t = \psi(D(z_t, t), \varepsilon_t, t) = d$ , where  $d$  is a possible value of  $D_t$  and  $\psi \in \Psi_t$ .*

The random variable  $Y_{t,j}^\psi(d)$  depends in part on future policy shocks such as  $\varepsilon_{t+j-1}$ , that is, random shocks that occur between time  $t$  and  $t + j$ . When we imagine changing  $d$  or  $\psi$  to generate potential

outcomes, the sequence of intervening shocks is held fixed. This is consistent with the tradition of impulse response analysis in macroeconomics. Our setup is more general, however, in that it allows the distributional properties of  $Y_{t,j}^\psi(d)$  to depend on the policy parameter  $d$  in arbitrary ways. In contrast, traditional impulse response analysis looks at the effect of  $d$  on the mean of  $Y_{t,j}^\psi(d)$  only.<sup>6</sup>

It also bears emphasizing that both the timing of policy adoption and the horizon matter for  $Y_{t,j}^\psi(d)$ . For example,  $Y_{t,j}^\psi(d)$  and  $Y_{t+1,j-1}^\psi(d)$  may differ even though both describe outcomes in period  $t+j$ . In particular,  $Y_{t,j}^\psi(d)$  and  $Y_{t+1,j-1}^\psi(d)$  may differ because  $Y_{t,j}^\psi(d)$  measures the effect of a policy change at time  $t$  on the outcome in time  $t+j$  and  $Y_{t+1,j-1}^\psi(d)$  measures the effect of period  $t+1$  policy on an outcome at time  $t+j$ .

Under the null hypothesis of no causal effect, potential and realized outcomes coincide. This is formalized in the next definition.

**Condition 1** *The sharp null hypothesis of no causal effects means that  $Y_{t,j}^{\psi'}(d') = Y_{t,j}^\psi(d)$ ,  $j > 0$  for all  $d, d'$  and for all possible policy functions  $\psi, \psi' \in \Psi_t$ . In addition, under the no-effects null hypothesis,  $Y_{t,j}^\psi(d) = Y_{t+j}$  for all  $d, \psi, t, j$ .*

In the simple situation studied by Rubin (1974), the no-effects null hypothesis states that  $Y_{0t} = Y_{1t}$ .<sup>7</sup> Our approach to causality testing leaves  $Y_{t,j}^\psi(d)$  unspecified. In contrast, it is common practice

---

<sup>6</sup>White (2006, 2009) develops a potential outcomes model for causal effects in a dynamic context. In contrast to our approach, White is concerned with the causal effect of policy sequences rather than individual policy shocks. White also discusses estimation of policy effects (as opposed to a focus on testing), but imposes stronger assumptions on the model relating outcomes and policies than we do.

<sup>7</sup>In a study of sequential randomized trials, Robins, Greenland and Hu (1999) define potential outcome  $Y_t^{(0)}$  as the outcome that would be observed in the absence of *any* current and past interventions, i.e. when  $D_t = D_{t-1} = \dots = 0$ . They

in econometrics to model the joint distribution of the vector of outcomes and policy variables ( $\chi_t$ ) as a function of lagged and exogenous variables or innovations in variables. It is therefore worth thinking about what potential outcomes would be in this case.

We begin with an example based on Bernanke and Blinder's (1992) SVAR model of monetary transmission (see also Bernanke and Mihov (1998)). This example illustrates how potential outcomes can be computed explicitly in simple linear models, and the link between observed and potential outcomes under the no-effects null.

**Example 1** *The economic environment is described by an SVAR of the form  $\Gamma_0 \chi_t = -\Gamma(L) \chi_t + (\eta'_t, \varepsilon_t)'$  where  $\Gamma_0$  is a matrix of constants conformable to  $\chi_t$  and  $\Gamma(L) = \Gamma_1 L + \dots + \Gamma_p L^p$  is a lag polynomial such that  $C(L) \equiv (\Gamma_0 + \Gamma(L))^{-1} = \sum_{k=0}^{\infty} C_k L^k$  exists. The policy innovations are denoted by  $\varepsilon_t$  and other structural innovations are  $\eta_t$ . Then,  $\chi_t = C(L) (\eta'_t, \varepsilon_t)'$  such that  $Y_t$  has a moving average representation*

$$Y_t = \sum_{k=0}^{\infty} c_{y\varepsilon,k} \varepsilon_{t-k} + \sum_{k=0}^{\infty} c_{y\eta,k} \eta_{t-k}$$

where  $c_{y\varepsilon,k}$  and  $c_{y\eta,k}$  are blocks of  $C_k$  partitioned conformably to  $Y_t$ ,  $\varepsilon_t$  and  $\eta_t$ . In this setup, potential outcomes are defined as

$$Y_{t,j}^{\psi}(d) = \sum_{k=0, k \neq j}^{\infty} c_{y\varepsilon,k} \varepsilon_{t+j-k} + \sum_{k=0}^{\infty} c_{y\eta,k} \eta_{t+j-k} + c_{y\varepsilon,j} d.$$

*These potential outcomes answer the following question: assume that everything else equal, which in this case means keeping  $\varepsilon_{t+j-k}$  and  $\eta_{t+j-k}$  fixed for  $k \neq j$ , how would the outcome variable  $Y_{t+j}$  change if we change the policy innovation from  $\varepsilon_t$  to  $d$ ? The sharp null hypothesis of no causal effect holds if and only*

---

denote by  $Y_t^{(1)}$  the set of values that could have potentially been observed if for all  $i \geq 0$ ,  $D_{t-i} = 1$ . This approach seems too restrictive to fit the macroeconomic policy experiments we have in mind.

if  $c_{y\varepsilon,j} = 0$  for all  $j$ . This is the familiar restriction that the impulse response function be identically equal to zero.<sup>8</sup>

When economic theory provides a model for  $\chi_t$ , as is the case for DSGE models, there is a direct relationship between potential outcomes and the solution of the model. As in Blanchard and Kahn (1980) or Sims (2001) a solution  $\tilde{\chi}_t = \tilde{\chi}_t(\bar{\varepsilon}_t, \bar{\eta}_t)$  is a representation of  $\chi_t$  as a function of past structural innovations  $\bar{\varepsilon}_t = (\varepsilon_t, \varepsilon_{t-1}, \dots)$  in the policy function and structural innovations  $\bar{\eta}_t = (\eta_t, \eta_{t-1}, \dots)$  in the rest of the economy. Further assuming that  $\psi(D(z_t, t), \varepsilon_t, t) = d$  can be solved for  $\varepsilon_t$  such that for some function  $\psi^*$ ,  $\varepsilon_t = \psi^*(D(z_t, t), d, t)$  we can then partition  $\tilde{\chi}_t = (\tilde{Y}_t, \tilde{X}_t, \tilde{D}_t)$  and focus on  $\tilde{Y}_t = \tilde{Y}_t(\bar{\varepsilon}_t, \bar{\eta}_t)$ . The potential outcome  $Y_{t,j}^\psi(d)$  can now be written as  $Y_{t,j}^\psi(d) = \tilde{Y}_{t+j}(\varepsilon_{t+j}, \dots, \varepsilon_{t+1}, \psi^*(\tilde{D}_t, d, t), \bar{\varepsilon}_{t-1}, \bar{\eta}_t)$ <sup>9</sup>. It is worth pointing out that the solution  $\tilde{\chi}_t$ , and thus the potential outcome  $Y_{t,j}^\psi(d)$ , in general both depends on  $D(\cdot, \cdot)$  and on the distribution of  $\varepsilon_t$ . With linear models, a closed form for  $\tilde{\chi}_t$  can be derived. Given such a functional relationship,  $Y_{t,j}^\psi(d)$  can be computed in an obvious way.<sup>10</sup>

Definition 1 extends the conventional potential outcome framework in a number of important ways. A

---

<sup>8</sup>In this example,  $Y_{t+1,j-1}^\psi(d)$  typically differs from  $Y_{t,j}^\psi(d)$ , except under the null hypothesis of no causal effects.

<sup>9</sup>When  $D_t = D(z_t, t) + \varepsilon_t$ ,  $\psi^*(D(z_t, t), d) = d - D(z_t, t)$ . However, the function  $\psi^*$  may not always exist. Then, it may be more convenient to index potential outcomes directly as functions of  $\varepsilon_t$  rather than  $d$ . In that case, one could define  $Y_{t,j}^\psi(e) = \tilde{Y}_{t+j}(\varepsilon_{t+j}, \dots, \varepsilon_{t+1}, e, \bar{\varepsilon}_{t-1}, \bar{\eta}_t)$  where we use  $e$  instead of  $d$  to emphasize the difference in definition. This distinction does not matter for our purposes and we focus on  $Y_{t,j}^\psi(d)$ .

<sup>10</sup>New Keynesian monetary models have multiple equilibria under certain interest rate targeting rules. Lubik and Schorfheide (2003) provide an algorithm to compute potential outcomes for linear rational expectations models with multiple equilibria. Multiplicity of equilibria is compatible with Condition 1 as long as the multiplicity disappears under the null hypothesis of no causal effects. Moreover, uniqueness of equilibria under the no-effects null need hold only for the component  $\tilde{Y}_t(\bar{\varepsilon}_t, \bar{\eta}_t)$  of  $\tilde{\chi}_t = (\tilde{Y}_t, \tilde{X}_t, \tilde{D}_t)$ .

key assumption in the cross-sectional causal framework is non-interference between units, or what Rubin (1978) calls the Stable Unit Treatment Value Assumption (SUTVA). Thus, in a cross-sectional context, the treatment received by one subject is assumed to have no causal effect on the outcomes of others. The overall proportion treated is also taken to be irrelevant. For a number of reasons, SUTVA may fail in a time series setup. First, because the units in a time series context are serially correlated, current outcomes depend on past policies. This problem is accounted for here by conditioning on the history of observed policies, covariates and outcomes, so that in practice potential outcomes reference alternative states of the world that might be realized for a given history. Second, and more importantly, since the outcomes of interest are often assumed to be equilibrium values, potential outcomes may depend on the distribution – and hence all possible realizations – of the unobserved component of policy decisions,  $\varepsilon_t$ . The dependence of potential outcomes on the distribution of  $\varepsilon_t$  is captured by  $\psi$ . Finally, the fact that potential outcomes depend on  $\psi$  allows them to depend directly on the decision-making rule used by policy makers even when policy realizations are fixed. Potential outcomes can therefore be defined in a rational-expectations framework where both the distribution of shocks and policy makers’ reaction to these shocks matter.

The framework up to this point defines causal effect in terms of unrealized potential or counterfactual outcomes. In practice, of course, we obtain only one realization each period, and therefore cannot directly test the non-causality null. Our tests therefore rely on the identification condition below, referred to in the cross-section treatment effects literature as “ignorability” or “selection-on-observables.” This condition allows us to establish a link between potential outcomes and the distribution of observed random variables.

**Condition 2** *Selection on observables:*

$$Y_{t,1}^\psi(d), Y_{t,2}^\psi(d), \dots \perp D_t | z_t, \text{ for all } d \text{ and } \psi \in \Psi_t.$$

The selection on observable assumption says that policies are independent of potential outcomes after appropriate conditioning. Note also that Condition 2 implies that  $Y_{t,1}^\psi(d), Y_{t,2}^\psi(d), \dots \perp \varepsilon_t | z_t$ . This is because  $D_t = \psi(z_t, \varepsilon_t, t)$  such that conditional on  $z_t$ , randomness in  $D_t$  is due exclusively to randomness in  $\varepsilon_t$ . We think of  $\varepsilon_t$  as shorthand for idiosyncratic factors such as those detailed for monetary policy by Romer and Romer (2004). These factors include the variation over time in policy makers' beliefs about the workings of the economy, decision-makers' tastes and goals, political factors, the temporary pursuit of objectives other than changes in the outcomes of interest (e.g., monetary policy that targets exchange rates instead of inflation or unemployment), and harder-to-quantify factors such as the mood and character of decision-makers. Conditional on observables, this idiosyncratic variation is taken to be independent of potential future outcomes.

The sharp null hypothesis in Condition 1 implies  $Y_{t,j}^{\psi'}(d') = Y_{t,j}^\psi(d) = Y_{t+j}$ . Substituting observed for potential outcomes in Condition 2 produces the key testable conditional independence assumption:

$$Y_{t+1}, \dots, Y_{t+j}, \dots \perp D_t | z_t. \tag{1}$$

In other words, conditional on observed covariates and lagged outcomes, there should be no relationship between treatment and outcomes.

Condition 2 plays a central role in the applied literature on testing the effects of monetary policy. For example, Bernanke and Blinder (1992), Gordon and Leeper (1994), Christiano, Eichenbaum and Evans (1996, 1999), and Bernanke and Mihov (1998) assume a block recursive structure to identify policy



shocks. In terms of Example 1, this is equivalent to imposing zero restrictions on the coefficients in  $\Gamma_0$  corresponding to the policy variables  $D_t$  in the equations for  $Y_t$  and  $X_t$  (see Bernanke and Mihov, 1998, p 874). Together with the assumption that  $\varepsilon_t$  and  $\eta_t$  are independent of each other and over time this implies Condition 2. To see this, note that conditional on  $z_t$ , the distribution of  $D_t$  depends only on  $\varepsilon_t$ , which is independent of the history of shocks that determine potential outcomes. Christiano, Eichenbaum and Evans (1999) discuss a variety of SVAR specifications that use recursive identification. The key assumption here is that an instantaneous response of conditioning variables to policy shocks can be ruled out a priori.

Tests based on Equation (1) can be seen as testing a restriction similar to the generalized version of Sims causality introduced by Chamberlain (1982). A natural question is how this relates to the Granger causality tests widely used in empirical work. Note that if  $X_t$  can be subsumed into the vector  $Y_t$ , Sims non-causality simplifies to  $Y_{t+1}, \dots, Y_{t+k}, \dots \perp D_t | \bar{Y}_t, \bar{D}_{t-1}$ . Chamberlain (1982) and Florens and Mouchart (1982, 1985) show that under plausible regularity conditions this is equivalent to generalized Granger non-causality, i.e.,

$$Y_{t+1} \perp D_t, \bar{D}_{t-1} | \bar{Y}_t. \quad (2)$$

In the more general case, however,  $D_t$  potentially causes  $X_{t+1}$ , so  $\bar{X}_t$  can not be subsumed into  $\bar{Y}_t$ .

Therefore, (1) does not imply

$$Y_{t+1} \perp D_t, \bar{D}_{t-1} | \bar{X}_t, \bar{Y}_t. \quad (3)$$

The fact that Sims and Granger causality are not generally equivalent was shown for the case of linear processes by Dufour and Tessier (1993).<sup>11</sup> We summarize the non-equivalence of Sims and Granger

---

<sup>11</sup>The relationship between Granger and Sims-type conditional independence restrictions is also discussed by Dufour and

causality in the following theorem:

**Theorem 1** *Let  $\chi_t$  be a stochastic process defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  as before, assuming also that conditional probability measures  $\Pr(Y_{t+1}, D_t | z_t)$  are well defined  $\forall t$  except possibly on a set of measure zero. Then (1) does not imply (3) and (3) does not imply (1).*

The intuition for the Granger/Sims distinction is that while Sims causality looks forward only at outcomes, the Granger causality relation is defined by conditioning on potentially endogenous responses to policy shocks and other disturbances. Even if the conditional independence assumption holds, the Granger test can be systematically misleading for the same reason that control for endogenous variables (i.e., other outcomes) complicates any kind of causal inference.<sup>12</sup>

Is the distinction between Granger and Sims causality empirically relevant in the money and income context? In research on monetary policy, Shapiro (1994) and Leeper (1997) argue that lagged inflation should be in the conditioning set when attempting to isolate the causal effect of monetary policy innovations. Suppose  $y_t$  is output,  $x_t$  is inflation, and  $D_t$  is a proxy for monetary policy. Suppose also that inflation is the only reason money affects output. In this case, Granger tests may fail to detect a causal link between monetary policy and output while Sims tests should detect this relationship. One way to understand this difference is through the impulse response function, which shows that Sims looks for an effect of structural innovations in policy (i.e.,  $\varepsilon_{D_t}$ ). In contrast, Granger non-causality is formulated as a restriction on the relation between output and all lagged variables, including covariates like inflation

---

Renault (1998) who consider a multi-step forward version of Granger causality testing, and Robins, Greenland, and Hu (1999) who state something like Theorem 1 without proof. Robins, Greenland and Hu also present restrictions on the joint process of  $w_t$  under which (1) implies (3) but these assumptions seem unrealistic for applications in macroeconomics.

<sup>12</sup>See, e.g., Section 3.2.3 in Angrist and Pischke (2009), on "bad control".

that themselves have responded to the policy shock of interest. Granger causality tests therefore give the wrong answer to a question that Sims causality tests answer correctly: will output change in response to a random manipulation of monetary policy?

The nonequivalence between Granger and Sims causality has important operational consequences: testing for (3) can be done easily with regression analysis by regressing  $Y_{t+1}$  on lags of  $D_t, Y_t$  and  $X_t$ , at least when additional distributional assumptions are imposed. While some implications of (1) can also be tested relatively easily with parametric models, testing the precise form of (1) is difficult unless  $D_t, Y_t$  and  $X_t$  can be nested within a linear dynamic model such as an SVAR model.<sup>13</sup> One of the main contributions of this paper is to relax linearity assumptions implicitly imposed on  $Y_{t,j}^\psi(d)$  by SVAR or regression analysis and to allow for non-linearities in the policy function.

In the remainder of the paper, we assume the policy variable of interest is multinomial. This is in the spirit of research focusing on Federal Reserve decisions regarding changes in the federal funds rate, which are by nature discrete (e.g., Hamilton and Jorda (2002)). Typically changes come in widely-publicized movements up or down, usually in multiples of 25 basis points if nonzero. As noted in Romer and Romer (2004), the Federal Reserve actively sets interest rate targets for most of the period since 1969, even when targeting was not as explicit as it is today. The discrete nature of monetary policy decisions leads naturally to a focus on the propensity-score, the conditional probability of a rate change (or a change of a certain magnitude or sign).<sup>14</sup>

---

<sup>13</sup>One example of a simple parametric test of (1) are our parametric Sims tests discussed in Section 5.

<sup>14</sup>Much of the empirical literature on the effects of monetary policy has focused on developing policy models for the federal funds rate. See, e.g., Bernanke and Blinder (1992), Christiano, Eichenbaum, and Evans (1996), and Romer and Romer (2004). In future work, we hope to develop an extension for continuous causal variables.

Under the non-causality null hypothesis it follows that  $\Pr(D_t|z_t, Y_{t+1}, \dots, Y_{t+j}, \dots) = \Pr(D_t|z_t)$ . A Sims-type test of the null hypothesis can therefore be obtained by augmenting the policy function  $p(z_t, \theta_0)$  with future outcome variables. This test has correct size though it will not have power against all alternatives. Below, we explore simple parametric Sims-type tests constructed by augmenting the policy function with future outcomes. But our main objective is use of the propensity score to develop a flexible class of semiparametric conditional independence tests that can be used to direct power in specific directions or to construct tests with power against general alternatives.

A natural substantive question at this point is what should go in the conditioning set for the policy propensity score and how this should be modeled. In practice, Fed policy is commonly modeled as being driven by a few observed variables like inflation and lagged output growth. Examples include Romer and Romer (1989, 2000, 2004) and others inspired by their work.<sup>15</sup> The fact that  $D_t$  is multinomial in our application also suggests that Multinomial Logit and Probit or similar models provide a natural functional form. A motivating example that seems especially relevant in this context is Shapiro (1994), who develops a parsimonious Probit model of Fed decision-making as a function of net present value measures of inflation and unemployment.<sup>16</sup> Importantly, while it is impossible to know for sure whether a given set of conditioning variables is adequate, our framework suggests a simple diagnostic test that can be used to decide when the model for the policy propensity score is consistent with the data.

---

<sup>15</sup>Stock and Watson (2002a, 2002b) propose the use of factor analysis to construct a low-dimensional predictor of inflation rates from a large dimensional data set. This approach has been used in the analysis of monetary policy by Bernanke and Boivin (2003) and Bernanke, Boivin and Elias (2005).

<sup>16</sup>Also related are Eichengreen, Watson and Grossman (1985), Hamilton and Jordà (2002), and Genberg and Gerlach (2004), who use ordered probit models for central bank interest rate targets.

### 3 Semiparametric Conditional Independence Tests Using the Propensity Score

We are interested in testing the conditional independence restriction  $y_t \perp D_t | z_t$ , where  $y_t$  takes values in  $\mathbb{R}^{k_1}$  and  $z_t$  takes values in  $\mathbb{R}^{k_2}$  with  $k_1 + k_2 = k$  finite. Typically,  $y_t = (Y'_{t+1}, \dots, Y'_{t+m})'$  but we can also focus on particular future outcomes, say,  $y_t = Y_{t+m}$ , when causal effects are thought to be delayed by  $m$  periods. Let  $v \in \mathbb{R}^k$  where  $v = (v'_1, v'_2)'$  is partitioned conformingly with  $(y'_t, z'_t)'$ . We assume that  $D_t$  is a discrete variable taking on  $\mathcal{M} + 1$  distinct values. Because  $\sum_{i=0}^{\mathcal{M}} \mathbf{1}(D_t = i) = 1$  and  $\sum_{i=0}^{\mathcal{M}} \Pr(D_t = i | z_t) = 1$  the conditional independence hypothesis can be written as a collection of the  $\mathcal{M}$  non-redundant moment conditions

$$\Pr(y_t \leq v_1, D_t = i | z_t) = \Pr(y_t \leq v_1 | z_t) \Pr(D_t = i | z_t) \text{ for } i = \{1, \dots, \mathcal{M}\}. \quad (4)$$

We use the shorthand notation  $p_i(z_t) = \Pr(D_t = i | z_t)$  and assume that  $p_i(z_t) = p_i(z_t, \theta)$  is known up to a parameter  $\theta$ . This is the policy propensity score. We assume that  $p(z_t, \theta_0)$  does not depend on  $t$  (in practice  $z_t$  might include time dummies). In an SVAR framework,  $p(z_t, \theta_0)$  corresponds to the SVAR policy-determination equation. In the recursive identification schemes discussed earlier, this equation can be estimated separately from the system. Our method differs in two important respects: We do not assume a linear relationship between  $D_t$  and  $z_t$  and we do not need to model the elements of  $z_t$  as part of a bigger system of simultaneous equations. This increases robustness and saves degrees of freedom relative to a conventional SVAR analysis.

A convenient representation of the hypotheses we are testing can be obtained by noting that under

the null,

$$\Pr(y_t \leq v_1, D_t = i | z_t) - \Pr(y_t \leq v_1 | z_t) p_i(z_t) = E[\mathbf{1}(y_t \leq v_1) (\mathbf{1}(D_t = i) - p_i(z_t)) | z_t] = 0. \quad (5)$$

These moment conditions can be written compactly in vector notation. We define  $\mathcal{M} \times 1$  vectors  $\mathcal{D}_t = (\mathbf{1}(D_t = 1), \dots, \mathbf{1}(D_t = \mathcal{M}))'$  and

$$p(z_t) = (p_1(z_t), \dots, p_{\mathcal{M}}(z_t))'$$

so that the moment conditions (5) can be expressed as

$$E[\mathbf{1}(y_t \leq v_1) (\mathcal{D}_t - p(z_t)) | z_t] = 0. \quad (6)$$

This leads to a simple interpretation of test statistics as looking for a relation between policy innovations,  $\mathcal{D}_t - p(z_t)$ , and the distribution of future outcomes. Note also that, like the Hirano, Imbens and Ridder (2003) and Abadie (2005) propensity-score-weighted estimators and the Robins, Mark, and Newey's (1992) partially linear estimator, test statistics constructed from moment condition (5) work directly with the propensity score; no matching step or nonparametric smoothing is required once estimates of the score have been constructed.

We define  $U_t = (y'_t, z'_t)'$  so that (6) can be expressed in terms of a collection of unconditional moment conditions. Correct specification of the policy model implies that  $E[(\mathcal{D}_t - p(z_t)) | z_t] = 0$ . Thus, testing (6) is equivalent to testing the unconditional moment condition  $E[\mathbf{1}(U_t \leq v) (\mathcal{D}_t - p(z_t))] = 0$  over all possible values of  $v$ . Appendix A presents a more general class of tests based on general test-functions,  $\phi(U_t, v)$  and not just indicators. In our empirical application,  $D_t$  indicates situations where the Fed raises, lowers or leaves the interest rates unchanged. Note that in our framework, interest rate increases may have causal effects while other sorts of movements do not. This possibility is explored by looking at

individual moment conditions

$$E [\mathbf{1}(y_t \leq v_1) ((\mathcal{D}_t = i) - p_i(z_t)) | z_t] = 0$$

for specific choices of  $i$ .

An implication of (6) is that the average policy effect is zero

$$E [E [\mathbf{1}(y_t \leq v_1) (\mathcal{D}_t - p(z_t)) | z_t]] = E [\mathbf{1}(y_t \leq v_1) (\mathcal{D}_t - p(z_t))] = 0. \tag{7}$$

In practice, the unconditional moment restriction (7) may be of greater interest than full conditional independence. Tests based on an unconditional restriction may also be more powerful.

The framework outlined here produces a specification test for the policy model. In particular, if the specification of  $p(z_t)$  is correct, the conditional moment restriction  $E [(\mathcal{D}_t - p(z_t)) | z_t] = 0$  holds, implying

$$E [\mathbf{1}(z_t \leq v_2) (\mathcal{D}_t - p(z_t))] = 0. \tag{8}$$

We use tests based on (8) to validate the empirical specification of  $p(z_t)$ .

Equation (5) shows that the hypothesis of conditional independence, whether formulated directly or for conditional moments, is equivalent to a martingale difference sequence (MDS) hypothesis for a certain empirical process. The moment condition in (5) implies that for any fixed  $v_1$ ,  $\mathbf{1}(y_t \leq v_1) (\mathcal{D}_t - p(z_t))$  is a MDS. Our test is a joint test of whether the set of all processes indexed by  $v_1 \in \mathbb{R}^{k_1}$  have the MDS property. We call this a functional martingale difference hypothesis. The functional MDS hypothesis is an extension of Koul and Stute (1999). One way in which our more general null hypothesis differs from their MDS test is that the dimension  $k$  of  $v$  is at least 2 while the simple MDS hypothesis is formulated for scalar  $v$ .<sup>17</sup>

---

<sup>17</sup>Another important difference is that in our setup, the process  $\mathbf{1}(y_t \leq y) (\mathcal{D}_t - p(z_t))$  is not Markovian even under the null hypothesis. This implies that the Koul and Stute results do not apply directly for our case.

To move from population moment conditions to the sample, we start by defining the empirical process

$V_n(v)$

$$V_n(v) = n^{-1/2} \sum_{t=1}^n m(y_t, D_t, z_t, \theta_0; v)$$

with

$$m(y_t, D_t, z_t, \theta; v) = \mathbf{1}\{U_t \leq v\} [D_t - p(z_t, \theta)].$$

Under regularity conditions that include stationarity of the observed process, we show in an Auxiliary Appendix that  $V_n(v)$  converges weakly to a limiting mean-zero Gaussian process  $V(v)$  with covariance function  $\Gamma(v, \tau)$

$$\Gamma(v, \tau) = \lim_{n \rightarrow \infty} E[V_n(v)V_n(\tau)']$$

where  $v, \tau \in \mathbb{R}^k$ .<sup>18</sup> Using the fact that under the null,  $E[D_t|z_t, y_t] = E[D_t|z_t] = p(z_t)$  and partitioning  $u = (u'_1, u'_2)'$  with  $u_2 \in [-\infty, \infty]^{k_2}$ , we define  $H(v)$  such that

$$H(v) = \int_{-\infty}^v (\text{diag}(p(u_2)) - p(u_2)p(u_2)') dF_u(u) \quad (9)$$

where  $\text{diag}(p(u_2))$  is the diagonal matrix with diagonal elements  $p_i(z_t)$ ,  $F_u(u)$  is the cumulative marginal distribution function of  $U_t$ . It follows that  $\Gamma(v, \tau) = H(v \wedge \tau)$  where  $\wedge$  denotes the element by element

---

<sup>18</sup>It seems likely that stationarity can be relaxed to allow for some distributional heterogeneity over time. But unit root and trend nonstationarity cannot be handled in our framework because the martingale transformations in Section 4.1 rely on Gaussian limit distributions. Park and Phillips (2000) develop a powerful limiting theory for the binary choice model when the explanatory variables have a unit root. Hu and Phillips (2002a, 2002b) extend Park and Phillips to the multinomial choice case and apply it to the fed funds target rate. The question of how to adapt these results to the problem of conditional independence testing is left for future work.



minimum. Let  $\|m\|^2 = \text{tr}(mm')$  be the usual Euclidean norm of a vector  $m$ . The statistic  $V_n(v)$  can be used to test the null hypothesis of conditional independence by comparing the value of  $\text{KS} = \sup_v \|V_n(v)\|$  or

$$\text{VM} = \int \|V_n(v)\|^2 dF_u(v) \quad (10)$$

with the limiting distribution of these statistics under the null hypothesis.

Implementation of statistics based on  $V_n(v)$  requires a set of appropriate critical values. Construction of critical values is complicated by two factors affecting the limiting distribution of  $V_n(v)$ . One is the dependence of the limiting distribution of  $V_n(v)$  on  $F_u(v)$ , which induces data-dependent correlation in the process  $V_n(v)$ . Hence, the nuisance parameter  $\Gamma(v, \tau)$  appears in the limiting distribution. This is handled in two ways: first, critical values for the limiting distribution of  $V_n(v)$  are computed numerically conditional on the sample in a way that accounts for the covariance structure  $\Gamma(v, \tau)$ . We discuss this procedure in Section 4.3. An alternative to numerical computation is to transform  $V_n(v)$  to a standard Gaussian process on the  $k$ -dimensional unit cube, following Rosenblatt (1952). The advantage of this approach is that asymptotic critical values can be based on standardized tables that only depend on the dimension  $k$  and the function  $\phi$ , but not on the distribution of  $U_t$  and thus not on the sample. We discuss how to construct these tables numerically in Section 5.

The second factor that affects the limiting distribution of  $V_n(v)$  is the fact that the unknown parameter  $\theta$  needs to be estimated. We use the notation  $\hat{V}_n(v)$  to denote test statistics that are based on an estimate,  $\hat{\theta}$ . Section 4 discusses a martingale transform proposed by Khmaladze (1988, 1993) to remove the effect of variability in  $\hat{V}_n(v)$  stemming from estimation of  $\theta$ . The corrected test statistic then has the same limiting distribution as  $V_n(v)$ , and thus, in a second step, critical values that are valid for  $V_n(v)$  can be

used to carry out tests based on the transformed version of  $\hat{V}_n(v)$ .

The combined application of the Rosenblatt and Khmaladze transforms leads to an asymptotically pivotal test. Pivotal statistics have the practical advantage of comparability across data-sets because the critical values for these statistics are not data-dependent. In addition to these practical advantages, bootstrapped pivotal statistics usually promise an asymptotic refinement (see Hall, 1992).

## 4 Implementation

As a first step, let  $\hat{V}_n(v)$  denote the empirical process of interest where  $p(z_t, \theta)$  is replaced by  $p(z_t, \hat{\theta})$  and the estimator  $\hat{\theta}$  is assumed to satisfy the following asymptotic linearity property:

$$n^{1/2} (\hat{\theta} - \theta_0) = n^{-1/2} \sum_{t=1}^n l(D_t, z_t, \theta_0) + o_p(1). \quad (11)$$

More detailed assumptions for the propensity score model are contained in Conditions 7 and 8 in Appendix C. In our context,  $l(D_t, z_t, \theta)$  is the score for the maximum likelihood estimator of the propensity score model. To develop a structure that can be used to account for the variability in  $\hat{V}_n(v)$  induced by the estimation of  $\theta$ , define the function  $\bar{m}(v, \theta) = E[m(y_t, D_t, z_t, \theta; v)]$  and let

$$\dot{m}(v, \theta) = -\frac{\partial \bar{m}(v, \theta)}{\partial \theta'}.$$

It therefore follows that  $\hat{V}_n(v)$  can be approximated by  $V_n(v) - \dot{m}(v, \theta_0)n^{-1/2} \sum_{t=1}^n l(D_t, z_t, \theta_0)$ . The empirical process  $\hat{V}_n(v)$  converges to a limiting process  $\hat{V}(v)$  with covariance function

$$\hat{\Gamma}(v, \tau) = \Gamma(v, \tau) - \dot{m}(v, \theta_0)L(\theta_0)\dot{m}(\tau, \theta_0)',$$

with  $L(\theta_0) = E[l(D_t, z_t, \theta_0)l(D_t, z_t, \theta_0)']$  as shown in the Auxiliary Appendix. Next we turn to details of the transformations. Section 4.1 discusses a Khmaladze-type martingale transformation that corrects

$\hat{V}(v)$  for the effect of estimation of  $\theta$ . Section 4.2 then discusses the problem of obtaining asymptotically distribution free limits for the resulting process. This problem is straightforward when  $v$  is a scalar, but extensions to higher dimensions are somewhat more involved.

## 4.1 Khmaladze Transform

The object here is to define a linear operator  $T$  with the property that the transformed process,  $W(v) = T\hat{V}(v)$ , is a mean zero Gaussian process with covariance function  $\Gamma(v, \tau)$ . While  $\hat{V}(v)$  has a complicated data-dependent limiting distribution (because of the estimated  $\theta$ ), the transformed process  $W(v)$  has the same distribution as  $V(v)$  and can be handled more easily in statistical applications. Khmaladze (1981, 1988, 1993) introduced the operator  $T$  in a series of papers exploring limiting distributions of empirical processes with possibly parametric means.

When  $v \in \mathbb{R}$ , the Khmaladze transform can be given some intuition. First, note that  $V(v)$  has independent increments  $\Delta V(v) = V(v+\delta) - V(v)$  for any  $\delta > 0$ . On the other hand, because  $\hat{V}(v)$  depends on the limit of  $n^{-1/2} \sum_{t=1}^n l(D_t, z_t, \theta_0)$  this process does not have independent increments. Defining  $\mathcal{F}_v = \sigma(\hat{V}(s), s \leq v)$ , we can understand the Khmaladze transform as being based on the insight that, because  $\hat{V}(v)$  is a Gaussian process,  $\Delta W(v) = \Delta \hat{V}(v) - E(\Delta \hat{V}(v) | \mathcal{F}_v)$  has independent increments. The Khmaladze transform thus removes the conditional mean of the innovation  $\Delta \hat{V}$ . When  $v \in \mathbb{R}^k$  with  $k > 1$  as in our application, this simple construction cannot be trivially extended because increments of  $V(v)$  in different directions of  $v$  are no longer independent. As explained in Khmaladze (1988), careful specification of the conditioning set  $\mathcal{F}_v$  is necessary to overcome this problem.

Following Khmaladze (1993), let  $\{A_\lambda\}$  be a family of measurable subsets of  $[-\infty, \infty]^k$ , indexed by

$\lambda \in [-\infty, \infty]$  such that  $A_{-\infty} = \emptyset$ ,  $A_\infty = [-\infty, \infty]^k$ ,  $\lambda \leq \lambda' \implies A_\lambda \subset A_{\lambda'}$  and  $A_{\lambda'} \setminus A_\lambda \rightarrow \emptyset$  as  $\lambda' \downarrow \lambda$ .

Define the projection  $\pi_\lambda f(v) = \mathbf{1}\{v \in A_\lambda\} f(v)$  and  $\pi_\lambda^\perp = 1 - \pi_\lambda$  such that  $\pi_\lambda^\perp f(v) = \mathbf{1}\{v \notin A_\lambda\} f(v)$ .

We then define the inner product

$$\langle f(\cdot), g(\cdot) \rangle \equiv \int f(u)' dH(u) g(u) \quad (12)$$

and, for

$$\bar{l}(v, \theta) = (\text{diag}(p(v_2)) - p(v_2)p(v_2)')^{-1} \frac{\partial p(v_2, \theta)}{\partial \theta'},$$

define the matrix

$$C_\lambda = \langle \pi_\lambda^\perp \bar{l}(\cdot, \theta), \pi_\lambda^\perp \bar{l}(\cdot, \theta) \rangle = \int \pi_\lambda^\perp \bar{l}(u, \theta)' dH(u) \pi_\lambda^\perp \bar{l}(u, \theta). \quad (13)$$

We note that the process  $V(v)$  can be represented in terms of a vector of Gaussian processes  $b(v)$  with covariance function  $H(v \wedge \tau)$  as  $V(\mathbf{1}\{\cdot \leq v\}) = V(v) = \int \mathbf{1}\{u \leq v\} db(u)$  and similarly  $V(l(\cdot, \theta_0)) = \int l(u, \theta_0) db(u)$  such that  $\hat{V}(f) = V(f(\cdot)) - \langle f(\cdot), \bar{l}(\cdot, \theta_0) \rangle \Sigma_\theta^{-1} V(\bar{l}(\cdot, \theta_0)')$ . The transformed statistic  $W(v)$  is then given by

$$W(v) \equiv T\hat{V}(v) = \hat{V}(v) - \int \langle \mathbf{1}\{\cdot \leq v\}, d(\pi_\lambda \bar{l}(\cdot, \theta)) \rangle C_\lambda^{-1} \hat{V}(\pi_\lambda^\perp \bar{l}(\cdot, \theta)') \quad (14)$$

where  $d(\pi_\lambda \bar{l}(\cdot, \theta))$  is the total derivative of  $\pi_\lambda \bar{l}(\cdot, \theta)$  with respect to  $\lambda$ .

We show in the Auxiliary Appendix that the process  $W(v)$  is zero mean Gaussian and has covariance function  $\Gamma(v, \tau)$ .

The transform above differs from that in Khmaladze (1993) and Koul and Stute (1999) in that  $\bar{l}(v, \theta)$  is different from the optimal score function that determines the estimator  $\hat{\theta}$ . The reason is that here  $H(v)$  is not a conventional cumulative distribution function as in these papers. It should also be emphasized

that unlike Koul and Stute (1999), we make no conditional homoskedasticity assumptions.<sup>19</sup>

Khmaladze (1993, Lemma 2.5) shows that tests based on  $W(v)$  and  $V(v)$  have the same local power against a certain class of local alternatives which are orthogonal to the score process  $l(\cdot, \theta_0)$ . The reason for this result is that  $T$  is a norm preserving mapping (see Khmaladze, 1993, Lemmas 3.4 and 3.10). The fact that local power is unaffected by the transformation  $T$  also implies that the choice of  $\{A_\lambda\}$  has no consequence for local power as long as  $A_\lambda$  satisfies the regularity conditions outlined above.

To construct the test statistic proposed in the theoretical discussion we must deal with the fact that the transformation  $T$  is unknown and needs to be replaced by an estimator  $T_n$ . This is obtained by replacing  $p(u_2)$  with  $p(u_2, \hat{\theta})$ ,  $H(u)$  with  $\hat{H}_n(u)$ ,  $C_\lambda$  with  $\hat{C}_\lambda$  and  $\hat{V}$  with  $\hat{V}_n$  in (14). Then  $T_n$  can be written as

$$\hat{W}_n(v) \equiv T_n \hat{V}_n(v) = \hat{V}_n(v) - \int d \left( \int_{u \leq v} d\hat{H}_n(u) (\pi_\lambda \bar{l}(u, \theta)) \right) \hat{C}_\lambda^{-1} \hat{V}_n(\pi_\lambda^\perp \bar{l}(\cdot, \hat{\theta})') \quad (15)$$

with  $\hat{V}_n(\pi_\lambda^\perp \bar{l}(\cdot, \hat{\theta})') = n^{-1/2} \sum_{s=1}^n \pi_\lambda^\perp \bar{l}(U_s, \hat{\theta})' (\mathcal{D}_s - p(z_s, \hat{\theta}))$  and the empirical distribution  $\hat{H}_n(u)$  and  $\hat{C}_\lambda$  are defined in Appendix B.

The transformed test statistic depends on the choice of the sets  $A_\lambda$  although, as pointed out earlier, the choice of  $A_\lambda$  does not affect local power. Computational convenience thus becomes a key criterion in selecting  $A_\lambda$ . Here we focus on sets

$$A_\lambda = [-\infty, \lambda] \times [-\infty, \infty]^{k-1}, \quad (16)$$

---

<sup>19</sup>Stute, Thies and Zhu (1998) analyze a test of conditional mean specification in an independent sample allowing for heteroskedasticity by rescaling the equivalent of our  $m(y_t, D_t, z_t, \theta_0; v)$  by the conditional variance. Here the relevant conditional variance depends on the unknown parameter  $\theta$ . Instead of correcting  $m(y_t, D_t, z_t, \theta_0; v)$  we adjust the transformation  $T$  in the appropriate way.

which lead to test statistics with simple closed form expressions. Denote the first element of  $y_t$  by  $y_{1t}$ .

Then (15) can be expressed more explicitly as

$$\hat{W}_n(v) = \hat{V}_n(v) - n^{-1/2} \sum_{t=1}^n \left[ \mathbf{1}\{U_t \leq v\} \frac{\partial p(z_t, \hat{\theta})}{\partial \theta'} \hat{C}_{y_{1t}}^{-1} n^{-1} \sum_{s=1}^n \mathbf{1}\{y_{1s} > y_{1t}\} \bar{l}(U_s, \hat{\theta})' (\mathcal{D}_s - p(z_s, \hat{\theta})) \right] \quad (17)$$

In Theorem 2 of Appendix C we show that  $\hat{W}_n(v)$  converges weakly to  $W(v)$ . In the next section we show how a further transformation can be applied that leads to a distribution free limit for the test statistics.

## 4.2 Rosenblatt Transform

The implementation strategy discussed above has improved operational characteristics when the data are modified using a transformation proposed by Rosenblatt (1952). This transformation produces a multivariate distribution that is i.i.d on the  $k$ -dimensional unit cube, and therefore leads to a test that can be based on standardized tables. Let  $U_t = [U_{t1}, \dots, U_{tk}]$  and define the transformation  $w = T_R(v)$  component wise by  $w_1 = F_1(v_1) = \Pr(U_{t1} \leq v_1)$ ,  $w_2 = F_2(v_2|v_1) = \Pr(U_{t2} \leq v_2|U_{t1} = v_1)$ , ...,  $w_k = F_k(v_k|v_{k-1}, \dots, v_1)$  where  $F_k(v_k|v_{k-1}, \dots, v_1) = \Pr(U_{tk} \leq v_k|U_{tk-1} = v_{k-1}, \dots, U_{t1} = v_1)$ . The inverse  $v = T_R^{-1}(w)$  of this transformation is obtained recursively as  $v_1 = F_1^{-1}(w_1)$ ,

$$v_2 = F_2^{-1}(w_2|F_1^{-1}(w_1)), \dots$$

Rosenblatt (1952) shows that the random vector  $w_t = T_R(U_t)$  has a joint marginal distribution which is uniform and independent on  $[0, 1]^k$ .

Using the Rosenblatt transformation we define

$$m_w(w_t, D_t, \theta|v) = \mathbf{1}\{w_t \leq w\} [\mathcal{D}_t - p([T_R^{-1}(w_t)]_z, \theta)]$$

where  $w = T_R(v)$  and  $z_t = [T_R^{-1}(w_t)]_z$  denotes the components of  $T_R^{-1}$  corresponding to  $z_t$ .

The null hypothesis is now that  $E[\mathbf{1}\{w_t \leq w\} \mathcal{D}_t | z_t] = E[\mathbf{1}\{w_t \leq w\} | z_t] p(z_t, \theta)$ , or equivalently,

$$E[m_w(w_t, D_t | v) | z_t] = 0.$$

Also, the test statistic  $V_n(v)$  becomes the marked process

$$V_{w,n}(w) = n^{-1/2} \sum_{t=1}^n m_w(w_t, D_t, \theta | w).$$

Rosenblatt (1952) notes that tests using  $T_R$  are generally not invariant to the ordering of the vector  $w_t$  because  $T_R$  is not invariant under such permutations.<sup>2021</sup>

We denote by  $V_w(v)$  the limit of  $V_{w,n}(v)$  and by  $\hat{V}_w(v)$  the limit of  $\hat{V}_{w,n}(v)$  which is the process obtained by replacing  $\theta$  with  $\hat{\theta}$  in  $V_{w,n}(v)$ . Define the transform  $T_w \hat{V}_w(w)$  as before by<sup>22</sup>

$$W_w(w) \equiv T_w \hat{V}_w(w) = \hat{V}_w(w) - \int \langle \mathbf{1}\{. \leq w\}, d\pi_\lambda \bar{l}_w(., \theta) \rangle C_\lambda^{-1} \hat{V}_w(\pi_\lambda^\perp \bar{l}_w(., \theta)'). \quad (18)$$

Finally, to convert  $W_w(w)$  to a process which is asymptotically distribution free we apply a modified version of the final transformation proposed by Khmaladze (1988, p. 1512) to the process  $W(v)$ . In particular, using the notation  $W_w(\mathbf{1}\{. \leq w\}) = W_w(w)$  to emphasize the dependence of  $W$  on  $\mathbf{1}\{. \leq w\}$ , and defining

$$h_w(.) = \left( \text{diag} \left( p([T_R^{-1}(.)]_z) \right) - p([T_R^{-1}(.)]_z) p([T_R^{-1}(.)]_z)' \right)$$

it follows from the previous discussion that

$$B_w(w) = W_w(\mathbf{1}\{. \leq w\}) (h_w(.))^{-1/2}$$

---

<sup>20</sup>In the working paper (Angrist and Kuersteiner, 2004) we discuss ways to resolve the problem of the ordering in  $w_t$ .

<sup>21</sup>Of course, the general form of our test statistic also depends on the choice of  $\phi(., .)$ , as outlined in Appendix A. This sort of dependence on the details of implementation is a common feature of consistent specification tests. From a practical point of view it seems natural to fix  $\phi(., .)$  using judgements about features of the data where deviations from conditional independence are likely to be easiest to detect (e.g., moments). In contrast, the  $w_t$  ordering is inherently arbitrary

<sup>22</sup>For a more detailed derivation see Appendix B.

is a Gaussian process with covariance function  $w \wedge w'$ .

In practice,  $w_t = T_R(U_t)$  is unknown because  $T_R$  depends on unknown conditional distribution functions. In order to estimate  $T_R$  we introduce the kernel function  $K_k(x)$  where  $K_k(x)$  is a higher order kernel satisfying Conditions 10 in Appendix C. A simple way of constructing higher order kernels is given in Bierens (1987). For  $\omega \geq 2$  let  $K_k(x) = (2\pi)^{-k/2} \sum_{j=1}^{\omega} \theta_j |\sigma_j|^{-k} \exp(-1/2x'x/\sigma_j^2)$  with  $\sum_{j=1}^{\omega} \theta_j = 1$  and  $\sum_{j=1}^{\omega} \theta_j |\sigma_j|^{2\ell} = 0$  for  $\ell = 1, 2, \dots, \omega - 1$ . Let  $m_n = O(n^{-(1-\kappa)/2k})$  for some  $\kappa$  with  $0 < \kappa < 1$  be a bandwidth sequence and define

$$\begin{aligned} \hat{F}_1(x_1) &= n^{-1} \sum_{t=1}^n \mathbf{1}\{U_{t1} \leq x_1\} \\ &\quad \vdots \\ \hat{F}_k(x_k | x_{k-1}, \dots, x_1) &= \frac{n^{-1} \sum_{t=1}^n \mathbf{1}\{U_{tk} \leq x_k\} K_{k-1}((x_{k-} - U_{tk-})/m_n)}{n^{-1} \sum_{t=1}^n K_{k-1}((x_{k-} - U_{tk-})/m_n)} \end{aligned}$$

where  $x_{k-} = (x_{k-1}, \dots, x_1)'$  and  $U_{tk-} = (U_{tk-1}, \dots, U_{t1})'$ . An estimate  $\hat{w}_t$  of  $w_t$  is then obtained from the recursions

$$\begin{aligned} \hat{w}_{t1} &= \hat{F}_1(U_{t1}) \\ &\quad \vdots \\ \hat{w}_{tk} &= \hat{F}_k(U_{tk} | U_{tk-1}, \dots, U_{t1}). \end{aligned}$$

We define  $\hat{W}_{w,n}(w) = T_{w,n} \hat{V}_{w,n}(w)$  where  $T_{w,n}$  is the empirical version of the Khmaladze transform applied to the vector  $w_t$ . Let  $\hat{W}_{\hat{w},n}(w)$  denote the process  $\hat{W}_{w,n}(w)$  where  $w_t$  has been replaced with  $\hat{w}_t$ . For a detailed formulation of this statistic see Appendix B. An estimate of  $h_w(w)$  is defined as

$$\hat{h}_w(\cdot) = \left( \text{diag} \left( p(\cdot, \hat{\theta}) \right) - p(\cdot, \hat{\theta}) p(\cdot, \hat{\theta})' \right).$$



The empirical version of the transformed statistic is

$$\begin{aligned}\hat{B}_{\hat{w},n}(w) &= \hat{W}_{\hat{w},n} \left( \mathbf{1}\{\cdot \leq w\} \hat{h}_w(\cdot)^{-1/2} \right) \\ &= n^{-1/2} \sum_{t=1}^n \mathbf{1}\{\hat{w}_t \leq w\} \hat{h}(z_t)^{-1/2} \left[ D_t - p(z_t, \hat{\theta}) - \hat{A}_{n,t} \right]\end{aligned}\quad (19)$$

where  $\hat{A}_{n,s} = n^{-1} \sum_{t=1}^n \mathbf{1}\{\hat{w}_{t1} > \hat{w}_{s1}\} \frac{\partial p(z_s, \hat{\theta})}{\partial \theta'} \hat{C}_{\hat{w}_{1s}}^{-1} \bar{l}(z_t, \hat{\theta})' \left( D_t - p(z_t, \hat{\theta}) \right)$ . Finally, Theorem 3 in Appendix C formally establishes that the process  $\hat{B}_{\hat{w},n}(v)$  converges to a Gaussian process with covariance function equal to the uniform distribution on  $[0, 1]^k$ .

Note that the convergence rate of  $\hat{B}_{\hat{w},n}(v)$  to a limiting random variable does not depend on the dimension  $k$  or the bandwidth sequence  $m_n$ . Theorem 3 shows that  $\hat{B}_{\hat{w},n}(w) \Rightarrow B_w(w)$  where  $B_w(w)$  is a standard Gaussian process. The set  $\Upsilon_{[0,1]}$  is defined as  $\Upsilon_{[0,1]} = \{w \in \Upsilon_\varepsilon | w = \pi_x w\}$  where  $\Upsilon_\varepsilon$  is a compact subset of the interior of  $[0, 1]^k$  with volume  $1 - \varepsilon$  for some  $\varepsilon > 0$ ,  $\pi_x w = \mathbf{1}(w \in A_x) w$  for some fixed  $x \in \mathbb{R}$  and  $A_x$  is the set defined in (16). The restriction to  $\Upsilon_{[0,1]}$  is needed to avoid problems of invertibility of  $\hat{C}_w^{-1}$ . It thus follows that transformed versions of the VM and KS statistics converge to functionals of  $B_w(w)$ . These results can be stated formally as

$$\text{VM}_w = \int_{\Upsilon_{[0,1]}} \left\| \hat{B}_{\hat{w},n}(w) \right\|^2 dw \Rightarrow \int_{\Upsilon_{[0,1]}} \|B_w(w)\|^2 dw \quad (20)$$

and

$$\text{KS}_w = \sup_{w \in \Upsilon_{[0,1]}} \left\| \hat{B}_{\hat{w},n}(w) \right\| \Rightarrow \sup_{w \in \Upsilon_{[0,1]}} \|B_w(w)\|. \quad (21)$$

Here  $\text{VM}_w$  and  $\text{KS}_w$  are the VM and KS statistics after both the Khmaladze and Rosenblatt transforms have been applied to  $\hat{V}_n(v)$ . In practice the integral in (20) and the supremum in (21) can be computed over a discrete grid. The asymptotic representations (20) and (21) make it possible to use asymptotic statistical tables. For the purposes of the empirical application below, we computed critical values for

the VM statistic. These critical values depend only on the dimension  $k$  and are thus distribution free.

### 4.3 Bootstrap-Based Critical Values

In addition to tests using critical values computed using asymptotic formulas, we also experimented with bootstrap critical values for the raw statistic,  $\hat{V}_n(v)$ , and the transformed statistic,  $\hat{B}_{\hat{w},n}(w)$ . This provides a check on the asymptotic formulas and gives some independent evidence on the advantages of the transformed statistic. Also, because the transformed statistic has a distribution free limit, we can expect an asymptotic refinement: tests based on bootstrapped critical values for this statistic should have more accurate size than bootstrap tests using  $\hat{V}_n(v)$ .

Our implementation of the bootstrap is similar to a procedure by Chen and Fan (1999) and Hansen (1996), a version of the wild bootstrap called conditional monte carlo. This procedure seems especially well-suited to time series data since it provides a simple strategy to preserve dependent data structures under resampling. Following Mammen (1993), the wild bootstrap error distribution is constructed by sampling  $\varepsilon_{t,s}^*$  for  $s = 1, \dots, S$  bootstrap replications according to

$$\varepsilon_{t,s}^* = \varepsilon_{t,s}^{**}/\sqrt{2} + \left( (\varepsilon_{t,s}^{**})^2 - 1 \right) / 2 \quad (22)$$

where  $\varepsilon_{t,s}^{**} \sim N(0, 1)$  is independent of the sample. Let the moment condition underlying the transformed test statistic (19) be denoted by

$$m_{T,t}(v, \hat{\theta}) = \mathbf{1}\{\hat{w}_t \leq w\} \hat{h}(z_t)^{-1/2} \left[ D_t - p(z_t, \hat{\theta}) - \hat{A}_{n,t} \right]$$

and write

$$\hat{B}_{\hat{w},n;s}^*(w) = n^{-1/2} \sum_{t=1}^n \varepsilon_{t,s}^* \left( m_{T,t}(v, \hat{\theta}) - \bar{m}_{n,T}(v, \hat{\theta}) \right) \quad (23)$$

to denote the test statistic in a bootstrap replication, with  $\bar{m}_{n;T}(v, \hat{\theta}) = n^{-1} \sum_{t=1}^n m_{T,t}(v, \hat{\theta})$ . The distribution of  $\varepsilon_{t,s}^*$  induced by (22) guarantees that the first three empirical moments of  $m_{T,t}(v, \hat{\theta}) - \bar{m}_{n;T}(v, \hat{\theta})$  are preserved in bootstrap samples. Theorem 4 in Appendix C shows that the asymptotic distribution of  $\hat{B}_{\hat{w},n}(w)$  under the null hypothesis is the same as the asymptotic distribution of  $\hat{B}_{\hat{w},n}^*(w)$  conditional on the data. This implies that critical values for  $\hat{B}_{\hat{w},n}(w)$  can be computed as follows: 1) Draw  $s = 1, \dots, S$  samples  $\varepsilon_{1,s}^*, \dots, \varepsilon_{n,s}^*$  independently from the distribution (22); 2) compute  $VM_s = \int_{\Upsilon_{[0,1]}} \left\| \hat{B}_{\hat{w},n;s}^*(w) \right\|^2 dw$  for  $s = 1, \dots, S$ ; 3) obtain the desired empirical quantile from the distribution of  $VM_s$ ,  $s = 1, \dots, S$ . The empirical quantile then approximates the critical value for  $\int_{\Upsilon_{[0,1]}} \left\| \hat{B}_{\hat{w},n}(w) \right\|^2 dw$ .

Bootstrap critical values for the untransformed statistic are based in an equivalent way on  $S$  bootstrap samples of

$$\hat{V}_{n;s}^*(v) = n^{-1/2} \sum_{t=1}^n \varepsilon_{t,s}^* \left( m(y_t, D_t, z_t, \hat{\theta}; v) - \bar{m}_n(v, \hat{\theta}) \right) \quad (24)$$

where  $\bar{m}_n(v, \hat{\theta}) = n^{-1} \sum_{t=1}^n m(y_t, D_t, z_t, \hat{\theta}; v)$  and  $\varepsilon_{t,s}^*$  is generated in the same way as before.

## 5 Causal Effects of Monetary Policy Shocks Revisited

We use the machinery developed here to test for the effects of monetary policy with data from Romer and Romer (2004). The key monetary policy variable in this study is the change in the FOMC's intended federal funds rate. This rate is derived from the narrative record of FOMC meetings and internal Federal Reserve memos. The conditioning variables for selection-on-observables identification are derived from Federal Reserve forecasts of the growth rate of real GNP/GDP, the GNP/GDP deflator, and the unemployment rate, as well as a few contemporaneous variables and lags. The relevant forecasts were prepared by Federal Reserve researchers and are called Greenbook forecasts.

The key identifying assumption in this context is that conditional on Greenbook forecasts and a handful of other variables, including lagged policy variables, changes in the *intended* federal funds target rate are independent of potential outcomes (in this case, the monthly percent change in industrial production). The Romers' (2004) detailed economic and institutional analysis of the monetary policy-making process makes their data and framework an ideal candidate for an investigation of causal policy effects using the policy propensity score.<sup>23</sup> In much of the period since the mid-1970s, and especially in the Greenspan era, the FOMC targeted the funds rate explicitly. The Romers argue, however, that even in the pre-Greenspan era, when the FOMC targeted the funds rate less closely, the central bank's intentions can be read from the documentary record. Moreover, the information used by the FOMC to make policy decisions is now available to researchers. The propensity-score approach begins with a statistical model predicting the intended federal funds rate as a function of the publicly available information used by the FOMC.

The propensity-score approach contrasts with SVAR-type identification strategies of the sort used by (among others) Bernanke and Blinder (1992), Bernanke, Boivin and Elias (2005), Christiano, Eichenbaum, and Evans (1996), Cochrane (1994), Leeper, Sims and Zha (1996). In this work, identification turns on a fully-articulated model of the macro economy, as well as a reasonably good approximation of the policy-making process. One key difference between the propensity-score approach developed here and the SVAR literature is that in the latter, policy variables and covariates entering the policy equation may also be endogenous variables. Identifying assumptions about how policy innovations are transmitted

---

<sup>23</sup>Romer and Romer (2004) can be seen as a response to critiques of Romer and Romer (1989) by Leeper (1997) and Shapiro (1994). These critics argued that monetary policy is forward-looking in a way that induces omitted variables bias in the Romers' (1989) regressions.

are then required to disentangle the causal effects of monetary policy from other effects.

Our approach is closer in spirit to the recursive identification strategy used by Christiano, Eichenbaum, and Evans (1999), hereafter CEE. Like ours, the CEE study makes the central bank’s policy function a key element in an analysis of monetary policy effects. Important differences, however, are that CEE formulate a monetary policy equation in terms of the actual federal funds rate and non-borrowed reserves and that they include contemporaneous values of real GDP, the GDP deflator and commodity prices as covariates. These variables are determined in part by market forces and are therefore potentially endogenous. For example, Sims and Zha (2006) argue that monetary aggregates and the producer price index are endogenous because of an immediate effect of monetary policy shocks on producer prices. In contrast, the *intended* funds rate used here is determined by forecasts of market conditions based on predetermined variables, and is therefore sequentially exogenous by construction. Finally, the CEE approach is parametric and relies on linear models for both outcomes and policy variables.

The substantive identifying assumption in our framework (as in Romer and Romer, 2004) is that, conditional on the information used by the FOMC and now available to outside researchers (such as Greenbook forecasts), changes in the intended funds rate are essentially idiosyncratic or “as good as randomly assigned.” At the same time, we don’t really know how best to model the policy propensity score - even maintaining the set of covariates, lag length is uncertain, for example. We therefore experiment with variations on the Romers’ original specification. We also consider an alternative somewhat less institutionally grounded model based on a simple Taylor rule. Our Taylor specification is motivated by Rotemberg and Woodford (1997).

Our reanalysis of the Romer data uses a discretized version of changes in the intended federal funds

rate. Specifically, to allow for asymmetric policy effects while keeping the model parsimonious, we treat policy as having three values: up, down, or no change. The change in the intended federal funds rate is denoted by  $dff_t$ , and the discretized change by  $dDff_t$ . For 29% of the monthly observations in our data, the intended funds rate fell, for 32% it rose, and the rest of time the intended rate was unchanged.<sup>24</sup> Following Hamilton and Jorda (2002), we fit ordered probit models with  $dDff_t$  as the dependent variable; this can be motivated by a linear latent-index model of central banker intentions.

The first specification we report on, which we call the baseline Romer model (a), uses the variables from Romer and Romer’s (2004) policy model as controls, with the modifications that the lagged level of the intended funds rate is replaced by the lagged change in the intended federal funds rate and the unemployment level is replaced by the unemployment innovation.<sup>25</sup> Our modifications are motivated in part by a concern that the lagged intended rate and the unemployment level are nonstationary. In addition, the lagged change in the intended federal funds rate captures the fact that the FOMC often acts in a sequence of small steps. This results in higher predicted probabilities of a change in the same direction conditional on past changes. A modified specification, constructed by dropping regressors without significant effects, leads to the restricted Romer model (b). To allow for non-linear dynamic responses, the lag-quadratic Romer model (c) adds a quadratic function of past intended changes in the federal funds rate to the restricted Romer model (b). We also consider versions of (a)-(c) using a

---

<sup>24</sup>We use the data set available via the Romer and Romer (2004) AER posting. Our sample period starts in March 1969 and ends in December 1996. Data for estimation of the policy propensity score are organized by “meeting month”: only observations during months with Federal Open Market meetings are recorded. In the early part of the sample there are a few occasions when the committee met twice in a month. These instances are treated as separate observations.

<sup>25</sup>The unemployment innovation is the Romers’  $\tilde{u}_{m0}$ , the Greenbook forecast for the unemployment rate in the current quarter, minus the unemployment rate in the previous month.

discretized variable for the lagged change in the intended federal funds rate. Romer models with discrete baseline are labeled (d), (e), and (f).

As an alternative to the policy model based on Romer and Romer (2004) we consider a Taylor-type model similar to the one used by Rotemberg and Woodford (1997). The Taylor models have  $dDff_t$  as the dependent variable in an ordered Probit model, as before. The covariates in this case consist of two lags of  $dff_t$ , 9 lags of the growth rate of real GDP, and 9 lags of the monthly inflation rate.<sup>26</sup> This baseline Taylor specification is labeled model (g). We also consider a modification replacing  $dff_{t-2}$  with  $(dff_{t-1})^2$  to capture nonlinearities in the lag-quadratic Taylor model (h). Finally, we look at Taylor models with discrete baseline controls, replacing lags of  $dff_t$  with the corresponding lags of  $dDff_t$ . These versions of models (g) and (h) are labeled (i) and (j).

As a benchmark for our semiparametric analysis, we begin with parametric Sims-type causality tests. These are simple parametric tests of the null hypothesis of no causal effect of monetary policy shocks on outcome variables, constructed by augmenting ordered Probit models for the propensity score with future outcome variables. Under the null hypothesis of no causal effect, future outcome variables should have insignificant coefficients in the policy model. This is the essence of Equation (1) and Condition 2.

Tables 1a and 1b report results for parametric Sims tests for the effect of policy on industrial production. The table shows t-statistics and significance levels for the coefficient on the cumulated change in the log of the non-seasonally adjusted index of industrial production,  $cIP_{t+k}$ , up to three years ahead. More specifically, each row in Tables 1a and 1b corresponds to separately estimated augmented ordered

---

<sup>26</sup>Monthly GDP is interpolated from quarterly using a program developed by Mönch and Uhlig (2005). We thank Emanuel Mönch and Harald Uhlig for providing the code for this. The inflation rate is calculated as the change in the log of the seasonally unadjusted CPI of urban consumers, less food and energy.

probit models  $p((z_t, cIP_{t+k}), \theta)$  for values of  $k$  up to 12 quarter leads. The variables  $z_t$  are the covariates specified for models (a)-(j), as defined in Appendix D. The models with lagged  $dDff_t$  on the right hand side point to a significant response to a change in monetary policy at a 5% significance level at 8 or more quarters lead. This result is robust across the Romer models (d)-(f) and the Taylor models (i)-(j). There is also isolated evidence of a response (at the 10% level) at earlier leads using models (e)-(j) in Table 1b. For models with  $dff_t$  on the right hand side, the lag pattern is more mixed. The baseline and restricted Romer models (a),(b) and the lag-quadratic Taylor model (h) predict a response after 7 quarters, while the lag-quadratic Romer model (c) predicts a response after 8 quarters and the baseline Taylor model (g) predicts a response after 6 quarters. The lag-quadratic Taylor model (h) generates an isolated initial impact of the monetary policy shock, but this does not persist at longer horizons. Tests at the 10 percent level generally show earlier effects, 6-7 quarters out for the restricted and lag-quadratic Romer models (b) and (c).

While easy to implement, the parametric Sims-causality tests do not tell us about differences in the effects of rate increases and decreases, and may not detect nonlinearities in the relationship between policy and outcomes, or effects of policy on higher-order moments. The semiparametric tests developed in Sections 3 and 4 do all this in an internally consistent way without the need for an elaborate model of the response function. The semiparametric tests can also be used to explore possible misspecification of the propensity score. This is done by substituting  $\mathbf{1}\{\tilde{z}_{ti} \leq v_{2i}\}$  for  $\mathbf{1}\{z_t \leq v_2\}$  in (8) where  $\tilde{z}_{it}$  denotes all the covariates that appear in models (a) through (j).

The specification tests reported in Table 2 suggest the baseline Romer model (a) and modifications



(c) and (e) fit well.<sup>2728</sup> The Taylor models fit less well, with moment restrictions violated most notably for the innovation in the Greenbook forecast for the percentage change in GDP/GNP. This suggests that the Taylor models do not fully account for all information the Federal Reserve seems to rely on in its policy decisions. The Taylor models also generate some rejections of moment conditions related to lagged  $dDff_t$ , an indication that they do not fully account for the dynamic pattern of Federal Reserve policy actions. The Romer models appear to implicitly account for lagged real GDP growth and inflation, in spite of the fact that these variables are not included in the Romer propensity score.

We now turn to the semiparametric causality tests based on the unconditional moment conditions in (7). All p-values reported in Tables 3-5 are based on the VM statistic defined in (20). In the first implementation,  $\mathcal{D}_t$  is a bivariate vector containing dummy variables for an up or down movement in  $dDff_t$ . This amounts to a joint test of the overall effect of a monetary policy shock, analogous to the parametric tests in Tables 1a and 1b.

The first set of semiparametric test results are reported in Tables 3a and 3b. As in Tables 2a and 2b, these tables show p-values and starred significance levels.<sup>29</sup> These tests look simultaneously at the

---

<sup>27</sup>The specification tests reported here should not be interpreted as pretests. A fully data-dependent propensity score model selection and subsequent testing approach is beyond the scope of this paper. We check for robustness to model specification by reporting causality tests for all models considered here.

<sup>28</sup>We show only results based on the VM statistic defined in (20) to save space. Other results are similar and available on request.

<sup>29</sup>Causality tests using ASY (not reported) and BSK are generally similar, though ASY appears to reject slightly more often. This is particularly true for Romer model (d) in Table 3b and to a lesser extent for models (a) and (f). Size distortions may arise due to multicollinearity induced by discretizing the lagged  $dff_t$  variables in these specifications. At the same time BSK sometimes rejects where ASY does not, for example, in Models (i) and (j) in Table 3b. Bootstrap based critical values based on untransformed statistic and based on (24) tend to reject much less often in most models, indicating some

significance of up and down movements in a single test statistic, in a manner analogous to the parametric tests in Table 1.

The results in Table 3 show significant effects at the 5% level starting 10 quarters ahead. The baseline Taylor model (a) also generates significant effects as early as in quarter 7, using both asymptotic and bootstrap critical values. The lag-quadratic Taylor model (h), and the Taylor models with discrete baseline (i) and (j) also generate significant effects starting in quarter 8. The restricted and lag-quadratic Romer models, (b), (c), (e) and (f), generate the longest lag in policy effects at about 10 quarters, although the restricted and lag-quadratic Romer models with discrete baseline (e) and (f) also show weaker significance at the 10% level as early as 4 quarters ahead in the restricted model (e) and 3 quarters ahead in the lag-quadratic model (f).

We also considered the effects of positive and negative monetary shocks separately. The asymmetric tests again use moment condition (7), but the tests in this case are constructed from  $\mathcal{D}_t = dDffU_t$  indicating upward movements in the intended funds rate and  $\mathcal{D}_t = -dDffD_t$  indicating decreases in the intended funds rate. Ordered Probit models for the policy propensity score generate the conditional expectation of both  $dDffD_t$  and  $dDffU_t$ , and can therefore be used to construct the surprise variable at the core of our testing framework. The asymmetric results are shown only for models that do well in the undersizing for these tests, especially in light of the parametric tests in Tables 1a and 1b.

Critical values for the asymptotic distribution were obtained by randomly drawing the  $k$ -dimensional vector  $U_{t,i}^{**}$  from a multivariate independent uniform distribution. In addition we draw independently  $\varepsilon_{t,i}$  from an iid standard normal distribution. The sample size was set to  $n = 300$  and 100,000 replications were done. We then compute  $B_i^{**}(w) = n^{-1/2} \sum_{t=1}^n \varepsilon_t^* \mathbf{1}\{U_{t,i}^{**} \leq x\}$  for each replication sample  $i$ . The final step consists in forming the sum  $B_i^{**} = n^{-1} \sum_{t=1}^n \mathbf{1}\{U_{j,i}^{**} \in \Upsilon_{[0,1]}\} \|B_i^{**}(U_{j,i}^{**})\|^2$ . Asymptotic critical values are then obtained from the quantiles of the empirical distribution of  $B_i^{**}$ .

model specification tests in Table 2. These are the baseline and lag-quadratic Romer models (a), (c), and the restricted Romer model with discrete baseline (e) and the lag-quadratic Taylor model (h).

The picture that emerges from Table 4 is mostly one of insignificant responses to a surprise reduction in the intended Federal Funds rate. In particular, the only models to show a statistically significant response to a decrease at the 5% level are the baseline Romer model (a) and the lag-quadratic Romer model (c), where a response appears after 10 quarters. Results for Taylor model, (h), generate an isolated significant test two-and-a-half years out. There is a less significant (10% level) response in the lag-quadratic Romer model with discrete baseline (e) and the lag-quadratic Taylor model (h) at a 10-11 quarter lead as well.

The results in Table 5 contrast with those in Table 4, showing significant effects of an increase in the funds rate after 6 quarters for Romer specification (a) and after 3 quarters for Romer specification (e). Taylor specification (h) also shows a strongly significant effect somewhere between quarter 7 or 8. Models (a) and (h) generate a less significant early response at 4 and 5 quarters. Also in contrast with Table 4, some of the results in table 5 are significant at the 1% level.

The results in Table 5 shed some light on the findings in Tables 3a and 3b, which pool up and down policy changes. The pooled results suggest a more immediate response for the baseline Romer specification (a) than for the lag-quadratic Taylor specification (h). This is consistent with the results in Table 5, where Romer model (a) uncovers a more immediate response to interest rate increases with a particularly strong response at 7 quarters lead but generates less significant test results than the Taylor models at leads farther out.

## 6 Conclusions

This paper develops a causal framework for time series data. The foundation of our approach is an adaptation of the potential-outcomes and selection-on-observables ideas widely used in cross-sectional studies. This adaptation leads to a definition of causality similar to that proposed by Sims (1972). For models with covariates, Sims causality differs from Granger causality, which potentially confuses endogenous system dynamics with the causal effects of isolated policy actions. In contrast, Sims causality hones in on the effect of isolated policy shocks relative to a well-defined counterfactual baseline.

Causal inference in our framework is based on a multinomial model for the policy assignment mechanism, a model we call the policy propensity score. In particular, we develop a new semiparametric test of conditional independence that uses the policy propensity score. This procedure tests the selection-on-observables null hypothesis that lies at the heart of much of the empirical work on time series causal effects. A major advantage of our approach is that it does not require researchers to model the process determining the outcomes of interest. The resulting test has power against all alternatives but can be fine-tuned to look at specific questions, such as mean independence or a particular direction of causal response. Our testing framework can also be used to evaluate the specification of the policy propensity score.

Our approach is illustrated with a re-analysis of the data and policy model in Romer and Romer (2004) along with a simple Taylor model. Our findings point to a significant response to monetary policy shocks after about 7 quarters, while the Taylor model and a restricted Romer specification shows responses that take a little longer to develop. These results are broadly in line with those in Romer and Romer (2004), which reports the strongest response to a monetary shock after about 2 years with continued effects

for another year. Our results therefore highlight the robustness of the Romers' original findings. An investigation into the statistical significance of different responses to rate increases and decreases shows an early and significant response to rate increases without much of a response to rate decreases. This result has not featured in most previous discussions of the causal effects of monetary shocks.

In contrast with the Romers findings and those reported here, SVAR studies generally report more immediate responses to a monetary shock. For example, Christiano, Eichenbaum and Evans (1999) report a decline in real GDP two quarters after a policy shock with the impulse response function showing a 'hump' shaped pattern and a maximal decline one to one and half years after the shock. Sims and Zha (2006) also find a statistically significant decline of real GDP in response to a money supply shock with most of the effect occurring in the first year after the shock. SVAR analysis of Taylor-type monetary policy functions in Rotemberg and Woodford (1997) similarly generates a response of real GDP after 2 quarters and a rapidly declining hump shaped response. Thus, while SVAR findings similarly suggest that monetary policy matters, some of the early impact that crops up in the SVAR literature may be generated in part by the structural assumptions used to identify these models.

An important topic for future research is the estimation of causal effects in situation where our tests reject the null hypothesis of no causal effect. In work in progress, we are exploring estimation strategies using a propensity score framework. The resulting estimators are similar in spirit to propensity score estimators for cross-sectional causal effects. However, a complication relative to the cross-sectional literature is the dynamic nature of responses to a policy shock. We are developing simple strategies to summarize and do inference for these dynamics.

## A General Test Statistics

This Appendix shows how to extend the statistics from test functions  $\mathbf{1}\{U_t \leq v\}$  to general functions  $\phi(U_t, v)$ . The null hypothesis of conditional independence can be represented very generally in terms of moment conditions for functions of  $U_t$ . Let  $\phi(\cdot, \cdot) : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{H}$  be a function of  $U_t$  and some index  $v$  where  $\mathbb{H}$  is some set. Our development below allows for  $\phi(U_t, v)$  to be a  $\mathcal{M} \times \mathcal{M}$  matrix of functions of  $U_t$  and  $v$  such that  $\mathbb{H} = \mathbb{R}^{\mathcal{M}} \times \mathbb{R}^{\mathcal{M}}$ . However, it is often sufficient to consider the case where  $\phi(\cdot, \cdot)$  is scalar valued with  $\mathbb{H} = \mathbb{R}$ , a possibility that is also covered by our theory. Under the null we then have  $E[\phi(U_t, v)(\mathcal{D}_t - p(z_t))|z_t] = 0$ . Examples of functions  $\phi(\cdot, \cdot)$  are  $\phi(U_t, v) = \exp(iv'U_t)$  where  $i = \sqrt{-1}$ , as suggested by Bierens (1982) and Su and White (2003), or  $\phi(U_t, v) = \mathbf{1}\{U_t \leq v\}$ , the case considered in Section 3.

While omnibus tests can detect departures from the null in all directions this is associated with a loss in power and may not shed light on specific alternatives of interest. Additional tests of practical relevance therefore focus on specific alternatives. An example is the test of the moment condition  $E[y_t(\mathcal{D}_t - p(z_t))|z_t] = 0$  which is rejected if there is correlation between  $y_t$  and the policy innovation conditional on  $z_t$ . Such a test can be implemented by choosing  $\phi(U_t, v) = y_t \mathbf{1}\{z_t \leq v_2\}$ . Generalizations to the effects on higher moments can be handled similarly.

To specify the generalized tests we extend the definition of  $V_n(v) = n^{-1/2} \sum_{t=1}^n m(y_t, D_t, z_t, \theta_0; v)$  by setting

$$m(y_t, D_t, z_t, \theta; v) = \phi(U_t, v) [\mathcal{D}_t - p(z_t, \theta)].$$

It follows that

$$\Gamma(v, \tau) = \lim_{n \rightarrow \infty} E[V_n(v)V_n(\tau)'] = \int \phi(u, v)dH(u)\phi(u, \tau)'$$

where  $H(v)$  is defined in (9). The transformation  $T$  now is given by

$$W(v) \equiv T\hat{V}(v) = \hat{V}(v) - \int \langle \phi(\cdot, v)', d(\pi_\lambda \bar{l}(\cdot, \theta)) \rangle C_\lambda^{-1} \hat{V}(\pi_\lambda^\perp \bar{l}(\cdot, \theta)') \quad (25)$$

where  $C_\lambda$ ,  $\hat{V}(\cdot)$  and  $\pi_\lambda^\perp \bar{l}(\cdot, \theta)$  are defined as before. In the same way define an estimator  $T_n$  where

$$\hat{W}_n(v) \equiv T_n \hat{V}_n(v) = \hat{V}_n(v) - \int \left( \int \phi(u, v) d\hat{H}_n(u) d(\pi_\lambda \bar{l}(u, \theta)) \right) \hat{C}_\lambda^{-1} \hat{V}_n(\pi_\lambda^\perp \bar{l}(\cdot, \hat{\theta})') \quad (26)$$

with  $\hat{V}_n(\pi_\lambda^\perp \bar{l}(\cdot, \hat{\theta})')$  and  $\hat{H}_n(v)$  as in Section 4.1. For  $A_\lambda = [-\infty, \lambda] \times [-\infty, \infty]^{k-1}$  one obtains

$$\hat{W}_n(v) = \hat{V}_n(v) - n^{-1/2} \sum_{t=1}^n \left[ \phi(U_t, v) \frac{\partial p(z_t, \hat{\theta})}{\partial \theta'} \hat{C}_{y_{1t}}^{-1} n^{-1} \sum_{s=1}^n \mathbf{1}\{y_{1s} > y_{1t}\} \bar{l}(U_s, \hat{\theta})' \left( \mathcal{D}_s - p(z_s, \hat{\theta}) \right) \right] \quad (27)$$

The Rosenblatt transform for  $\hat{W}_n(v)$  based on general functions  $\phi(\cdot, \cdot)$  is obtained by extending (18) to

$$W_w(w) \equiv T_w \hat{V}_w(w) = \hat{V}_w(w) - \int \langle \phi(\cdot, w)', d\pi_\lambda \bar{l}_w(\cdot, \theta) \rangle C_\lambda^{-1} \hat{V}_w(\pi_\lambda^\perp \bar{l}_w(\cdot, \theta)'). \quad (28)$$

and

$$B_w(w) = W_w(\phi(\cdot, w)(h_w(\cdot))^{-1/2}).$$

is a Gaussian process with covariance function  $\int_0^1 \cdots \int_0^1 \phi(u, w)\phi(u, w')' du$ .

The empirical version of the transformed statistic is

$$\begin{aligned} \hat{B}_{\hat{w},n}(w) &= \hat{W}_{\hat{w},n}(\phi(\cdot, w)\hat{h}_w(\cdot)^{-1/2}) \\ &= n^{-1/2} \sum_{t=1}^n \phi(\hat{w}_t, w) \hat{h}(z_t)^{-1/2} \left[ D_t - p(z_t, \hat{\theta}) - \hat{A}_{n,t} \right] \end{aligned} \quad (29)$$

where  $\hat{A}_{n,s}$  is as defined before. For the bootstrapped statistic  $\hat{B}_{\hat{w},n;s}^*(w)$  replace  $m_{T,t}(v, \hat{\theta})$  with

$$m_{T,t}(v, \hat{\theta}) = \phi(\hat{w}_t, w) \hat{h}(z_t)^{-1/2} \left[ D_t - p(z_t, \hat{\theta}) - \hat{A}_{n,t} \right]$$

in (23).

## B Implementation Details

### B.1 Details for the Khmaladze Transform

To construct the test statistic proposed in the theoretical discussion we must deal with the fact that the transformation  $T$  is unknown and needs to be replaced by an estimator. In this section, we discuss the details that lead to the formulation in (17). We also present results for general sets  $A_\lambda$ . We start by defining the empirical distribution

$$\hat{F}_u(v) = n^{-1} \sum_{t=1}^n \mathbf{1}\{U_t \leq v\}, \quad (30)$$

and let

$$\begin{aligned} H_n(v) &= \int_{-\infty}^v (\text{diag}(p(u_2, \theta_0)) - p(u_2, \theta_0)p(u_2, \theta_0)') d\hat{F}_u(u) \\ &= n^{-1} \sum_{t=1}^n (\text{diag}(p(z_t, \theta_0)) - p(z_t, \theta_0)p(z_t, \theta_0)') \mathbf{1}\{U_t \leq v\} \end{aligned}$$

as well as

$$\begin{aligned} \hat{H}_n(v) &= \int_{-\infty}^v (\text{diag}(p(z_t, \hat{\theta})) - p(z_t, \hat{\theta})p(z_t, \hat{\theta})') d\hat{F}_u(u) \\ &= n^{-1} \sum_{t=1}^n (\text{diag}(p(z_t, \hat{\theta})) - p(z_t, \hat{\theta})p(z_t, \hat{\theta})') \mathbf{1}\{U_t \leq v\}. \end{aligned}$$

We now use the sets  $A_\lambda$  and projections  $\pi_\lambda$  as defined in Section 4.1. Let

$$\begin{aligned} \hat{C}_\lambda &= \int \pi_\lambda^\perp \bar{l}(v, \hat{\theta}) d\hat{H}_n(v) \pi_\lambda^\perp \bar{l}(v, \hat{\theta})' \\ &= n^{-1} \sum_{t=1}^n (1 - \mathbf{1}\{U_t \in A_\lambda\}) \bar{l}(U_t, \hat{\theta})' \left( \text{diag}(p(z_t, \hat{\theta})) - p(z_t, \hat{\theta})p(z_t, \hat{\theta})' \right) \bar{l}(U_t, \hat{\theta}) \end{aligned}$$

such that

$$T_n \hat{V}_n(v) = \hat{V}_n(v) - \int d \left( \int \phi(u, v) d\hat{H}_n(u) \pi_\lambda \bar{l}(u, \hat{\theta}) \right) \hat{C}_\lambda^{-1} \hat{V}_n(\pi_\lambda^\perp \bar{l}(u, \hat{\theta}))$$



where

$$\int \phi(u, v) d\hat{H}_n(u) \pi_\lambda \bar{l}(\cdot, \hat{\theta}) = n^{-1} \sum_{t=1}^n \mathbf{1}\{U_t \in A_\lambda\} \phi(U_t, v) \frac{\partial p(z_t, \hat{\theta})}{\partial \theta'}.$$

Finally, write

$$\hat{V}_n(\pi_\lambda^\perp \bar{l}(u, \hat{\theta})) = n^{-1/2} \sum_{t=1}^n (1 - \mathbf{1}\{U_t \in A_\lambda\}) \bar{l}(U_t, \hat{\theta})' (\mathcal{D}_t - p(z_t, \hat{\theta})).$$

We now specialize the choice of sets  $A_\lambda$  to  $A_\lambda = [-\infty, \lambda] \times [-\infty, \infty]^{k-1}$ . Denote the first element of  $y_t$  by  $y_{1t}$ . Then

$$\hat{C}_\lambda = n^{-1} \sum_{t=1}^n \mathbf{1}\{y_{1t} > \lambda\} \bar{l}(z_t, \hat{\theta}) \left( \text{diag} \left( p(z_t, \hat{\theta}) \right) - p(z_t, \hat{\theta}) p(z_t, \hat{\theta})' \right) \bar{l}(z_t, \hat{\theta})', \quad (31)$$

$$\hat{V}_n(\pi_\lambda^\perp \bar{l}(u, \hat{\theta})) = n^{-1/2} \sum_{t=1}^n \mathbf{1}\{y_{1t} > \lambda\} \bar{l}(U_t, \hat{\theta})' (\mathcal{D}_t - p(z_t, \hat{\theta})) \quad (32)$$

and

$$\int \phi(u, v) d\hat{H}_n(u) \pi_\lambda \bar{l}(u, \hat{\theta}) = n^{-1} \sum_{t=1}^n \mathbf{1}\{y_{1t} \leq \lambda\} \phi\{U_t, v\} \frac{\partial p(z_t, \hat{\theta})}{\partial \theta'} \quad (33)$$

Combining 31, 32 and 33 then leads to the formulation 17.

## B.2 Details for the Rosenblatt Transform

As before implementation requires replacement of  $\theta$  with an estimate. We therefore work with the process  $\hat{V}_{w,n}(v) = n^{-1/2} \sum_{t=1}^n m_w(w_t, \mathcal{D}_t, \hat{\theta}; w)$ . Define

$$E[m_w(w_t, D_t, \theta); w] = \int_0^1 \cdots \int_0^1 \phi(u, w) \left( p\left([T_R^{-1}(u)]_z, \theta_0\right) - p\left([T_R^{-1}(u)]_z, \theta\right) \right) du$$

such that  $\dot{m}(w, \theta)$  evaluated at the true parameter value  $\theta_0$  is

$$\begin{aligned}\dot{m}_w(w, \theta_0) &= E[\phi(U_t, w)\partial p(z_t, \theta_0)/\partial\theta'] \\ &= \int_{[0,1]^k} \phi(u, w) \frac{\partial p([T_R^{-1}(u)]_z, \theta_0)}{\partial\theta'} du\end{aligned}$$

It therefore follows that  $\hat{V}_{w,n}(v)$  can be approximated by  $V_{w,n}(v) - \dot{m}_w(w, \theta_0)'n^{-1/2} \sum_{t=1}^n l(\mathcal{D}_t, z_t, \theta_0)$ .

This approximation converges to a limiting process  $\hat{V}_w(v)$  with covariance function

$$\hat{\Gamma}_w(w, \tau) = \Gamma_w(w, \tau) - \dot{m}_w(w, \theta_0)'L(\theta_0)\dot{m}_w(\tau, \theta_0)$$

where

$$\Gamma_w(w, \tau) = \int_{[0,1]^k} \phi(u, w)h_w(u) \phi(u, \tau)'du.$$

where  $h_w(\cdot, \theta) = (\text{diag}(p([T_R^{-1}(\cdot), \theta]_z)) - p([T_R^{-1}(\cdot), \theta]_z)p([T_R^{-1}(\cdot)]_z, \theta)')$  and  $h_w(\cdot) \equiv h_w(\cdot, \theta_0)$ .

We represent  $\hat{V}_w$  in terms of  $V_w$ . Let  $V_w(l_w(\cdot, \theta_0)) = \int l_w(w, \theta_0)b_w(dv)$  where  $b_w(v)$  is a Gaussian process on  $[0, 1]^k$  with covariance function  $\Gamma_w(v, \tau)$  as before, for any function  $l_w(w, \theta)$ . Also, define

$$\bar{l}_w(w, \theta) = h_w(w, \theta)^{-1} \frac{\partial p([T_R^{-1}(w)]_z, \theta)}{\partial\theta'}$$

such that  $\hat{V}_w(w) = V_w(w) - \dot{m}_w(w, \theta_0)V_w(\bar{l}_w(w, \theta))$  as before.

Let  $\{A_{w,\lambda}\}$  be a family of measurable subsets of  $[0, 1]^k$ , indexed by  $\lambda \in [0, 1]$  such that  $A_{w,0} = \emptyset$ ,  $A_{w,1} = [0, 1]^k$ ,  $\lambda \leq \lambda' \implies A_{w,\lambda} \subset A_{w,\lambda'}$  and  $A_{w,\lambda'} \setminus A_{w,\lambda} \rightarrow \emptyset$  as  $\lambda' \downarrow \lambda$ . We then define the inner product  $\langle f(\cdot), g(\cdot) \rangle_w \equiv \int_{[0,1]^k} f(w)'dH_w(w)g(w)$  where

$$H_w(w) = \int_{u \leq w} h_w(u) du$$

and the matrix

$$C_{w,\lambda} = \langle \pi_\lambda^\perp \bar{l}_w(\cdot, \theta), \pi_\lambda^\perp \bar{l}_w(\cdot, \theta) \rangle_w = \int \pi_\lambda^\perp \bar{l}_w(w, \theta)'dH_w(w)\pi_\lambda^\perp \bar{l}_w(w, \theta).$$

and define the transform  $T_w V_w(w)$  as before by

$$W_w(w) \equiv T_w \hat{V}_w(w) = \hat{V}_w(w) - \int \langle \phi(\cdot, w)', d\pi_{\lambda} \bar{l}_w(\cdot, \theta) \rangle C_{\lambda}^{-1} \hat{V}_w(\pi_{\lambda}^{\perp} \bar{l}_w(\cdot, \theta)').$$

Finally, to convert  $W_w(w)$  to a process which is asymptotically distribution free we apply a modified version of the final transformation proposed by Khmaladze (1988, p. 1512) to the process  $W(w)$ . In particular, using the notation  $W_w(\phi(\cdot, w)) = W_w(w)$  to emphasize the dependence of  $W$  on  $\phi$ , it follows from the previous discussion that

$$B_w(w) = W_w(\phi(\cdot, w)(h_w(\cdot))^{-1/2})$$

where  $B_w(w)$  is a Gaussian process on  $[0, 1]^k$  with covariance function  $\int_0^1 \cdots \int_0^1 \phi(u, w)\phi(u, w') du$ .

The empirical version of  $W_w(w)$ , denoted by  $\hat{W}_{w,n}(w) = \hat{T}_w \hat{V}_{w,n}(w)$ , is obtained as before from

$$\hat{W}_{w,n}(w) = n^{-1/2} \sum_{t=1}^n \left[ m_w(w_t, D_t, \hat{\theta}|w) - \phi(w_t, w) \frac{\partial p(z_t, \hat{\theta})}{\partial \theta'} \hat{C}_{w_t}^{-1} n^{-1} \sum_{s=1}^n \mathbf{1}\{w_{s1} > w_{t1}\} \bar{l}(z_s, \hat{\theta})' (\mathcal{D}_s - p(z_s, \hat{\theta})) \right]$$

where  $\hat{C}_{w_{s1}} = n^{-1} \sum_{t=1}^n \mathbf{1}\{w_{t1} > w_{s1}\} \bar{l}(z_t, \hat{\theta})' h(z_t, \hat{\theta}) \bar{l}(z_t, \hat{\theta})$ .

## C Formal Results

This Appendix provides formal results on the distribution of the test statistics described above. Let  $\chi_t = [y_t', z_t', D_t]'$  be the vector of observations. Assume that  $\{\chi_t\}_{t=1}^{\infty}$  is strictly stationary with values in the measurable space  $(\mathbb{R}^{k+1}, \mathcal{B}^{k+1})$  where  $\mathcal{B}^{k+1}$  is the Borel  $\sigma$ -field on  $\mathbb{R}^{k+1}$  and  $k$  is fixed with  $2 \leq k < \infty$ . 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

$$\beta_m = \sup_{l \geq 1} E \left[ \sup_{A \in \mathcal{A}_{l+m}^{\infty}} |\Pr(A|\mathcal{A}_1^l) - \Pr(A)| \right] \rightarrow 0 \text{ as } m \rightarrow \infty. \quad (34)$$

**Condition 3** Let  $\chi_t$  be a stationary, absolutely regular process such that for some  $2 < p < \infty$  the  $\beta$ -mixing coefficient of  $\chi_t$  defined in (34) satisfies  $m^{(p+\delta)/(p-2)} (\log m)^{2(p-1)/(p-2)} \beta_m \rightarrow 0$  for some  $\delta > 0$ .

**Condition 4** Let  $F_u(u)$  be the marginal distribution of  $U_t$ . Assume that  $F_u(\cdot)$  is absolutely continuous with respect to Lebesgue measure on  $\mathbb{R}^k$  and has a density  $f_u(u)$  with  $f_u(u) > 0$  for all  $u \in \mathbb{R}^k$ .

**Condition 5** The matrix of functions  $\phi(\cdot, \cdot)$  belongs to a VC subgraph class of functions (see Pollard 1984) with envelope  $M(\chi_t)$  such that  $E \|M(\chi_t)\|^{p+\delta} < \infty$  for the same  $p$  and  $\delta$  as in Condition 3.

**Condition 6** Let  $H(v)$  be as defined in (9). Assume that  $H(v)$  is absolutely continuous in  $v$  with respect to Lebesgue measure and for all  $v, \tau$  such that  $v \leq \tau$  with  $v_i < \tau_i$  for at least one element  $v_i$  of  $v$  it follows that  $H(v) < H(\tau)$ . Let the  $\mathcal{M} \times \mathcal{M}$  matrix of derivatives  $h(v) = \partial^k H(v) / \partial v_1 \dots \partial v_k$  and assume that  $\det(h(v)) > 0$  for all  $v \in \mathbb{R}^k$ .

**Condition 7** Let  $\theta_0 \in \Theta$  where  $\Theta \subset \mathbb{R}^d$  is a compact set and  $d < \infty$ . Assume that  $E[D_t|z_t] = p(z_t|\theta_0)$  and for all  $\theta \neq \theta_0$  it follows  $E[D_t|z_t] \neq p(z_t|\theta)$ . Assume that  $p(z_t|\theta)$  is differentiable a.s. for  $\theta \in \{\theta \in \Theta \mid \|\theta - \theta_0\| \leq \delta\} \equiv N_\delta(\theta_0)$  for some  $\delta > 0$ . Let  $N(\theta_0)$  be a compact subset of the union of all neighborhoods  $N_\delta(\theta_0)$  where  $\partial p(z_t|\theta) / \partial \theta$ ,  $\partial^2 p(z_t|\theta) / \partial \theta_i \partial \theta_j$  exists and assume that  $N(\theta_0)$  is not empty. Let  $\partial p_i(z_t|\theta) / \partial \theta_j$  be the  $i, j$ -th element of the matrix of partial derivatives  $\partial p(z_t|\theta) / \partial \theta'$  and let  $\bar{l}_{i,j}(z_t, \theta)$  be the  $i, j$ -th element of  $\bar{l}(z_t, \theta)$ . Assume that there exists a function  $B(x)$  and a constant  $\alpha > 0$  such that

$$|\partial p_i(x|\theta) / \partial \theta_j - \partial p_i(x|\theta') / \partial \theta_j| \leq B(x) \|\theta - \theta'\|^\alpha,$$

$|\partial^2 p_k(x|\theta) / \partial \theta_i \partial \theta_j - \partial^2 p_k(x|\theta) / \partial \theta_i \partial \theta_j| \leq B(x) \|\theta - \theta'\|^\alpha$  and  $|\partial \bar{l}_{i,j}(x|\theta) / \partial \theta_k - \partial \bar{l}_{i,j}(x|\theta') / \partial \theta_k| \leq B(x) \|\theta - \theta'\|^\alpha$  for all  $i, j, k$  and  $\theta, \theta' \in \text{int } N(\theta_0)$ ,  $E|B(z_t)|^{2+\delta} < \infty$ ,  $E|\partial p_i(z_t|\theta_0) / \partial \theta_j|^{4+\delta} < \infty$ ,  $E[p_i(z_t, \theta_0)^{-(4+\delta)}] < \infty$  and  $E\left[|\partial p_i(z_t|\theta_0) / \partial \theta_j|^{\frac{4+\delta}{2}}\right] < \infty$  for all  $i, j$  and some  $\delta > 0$ .

**Condition 8** Let  $l(D_t, z_t, \theta) = \Sigma_\theta^{-1} \frac{\partial p'(z_t, \theta)}{\partial \theta} h(z_t, \theta)^{-1} (\mathcal{D}_t - p(z_t, \theta))$ ,

$$h(z_t, \theta) = (\text{diag}(p(z_t, \theta)) - p(z_t, \theta)p(z_t, \theta)')$$

and

$$\Sigma_\theta = E \left[ \frac{\partial p'(D_t | z_t, \theta)}{\partial \theta} h(z_t, \theta)^{-1} \frac{\partial p(D_t | z_t, \theta)}{\partial \theta'} \right]. \quad (35)$$

Assume that  $\Sigma_\theta$  is positive definite for all  $\theta$  in some neighborhood  $N \subset \Theta$  such that  $\theta_0 \in \text{int } N$  and  $0 < \|\Sigma_\theta\| < \infty$  for all  $\theta \in N$ . Let  $l_i(D_t, z_t, \theta)$  be the  $i$ -th element of  $l(D_t, z_t, \theta)$  defined in (11). Assume that there exists a function  $B(x_1, x_2)$  and a constant  $\alpha > 0$  such that  $\|\partial l_i(x_1, x_2, \theta) / \partial \theta_j - \partial l_i(x_1, x_2, \theta') / \partial \theta_j\| \leq B(x_1, x_2) \|\theta - \theta'\|^\alpha$  for all  $i$  and  $\theta, \theta' \in \text{int } N$ ,  $E[B(D_t z_t)] < \infty$  and  $E|l(D_t, z_t, \theta)| < \infty$  for all  $i$ .

**Condition 9** Let  $\{A_\lambda\}$  be a family of measurable subsets of  $[-\infty, \infty]^k$ , indexed by  $\lambda \in [-\infty, \infty]$  such that  $A_{-\infty} = \emptyset$ ,  $A_\infty = [-\infty, \infty]^k$ ,  $\lambda \leq \lambda' \implies A_\lambda \subset A_{\lambda'}$  and  $A_{\lambda'} \setminus A_\lambda \rightarrow \emptyset$  as  $\lambda' \downarrow \lambda$ . Assume that the sets  $\{A_\lambda\}$  form a V-C class (polynomial class) of sets as defined in Pollard (1984, p.17). Define  $\langle f(\cdot), g(\cdot) \rangle$  as in (12) and  $C_\lambda$  as in (13). Assume that  $\langle f(v), \pi_\lambda g(v) \rangle$  is absolutely continuous in  $\lambda$  and  $C_\lambda$  is invertible for  $\lambda \in [-\infty, \infty)$ .

**Condition 10** The density  $f_u(u)$  is continuously differentiable to some integral order  $\omega \geq \max(2, k)$  on  $\mathbb{R}^k$  with  $\sup_{x \in \mathbb{R}^k} |D^\mu f_u(x)| < \infty$  for all  $|\mu| \leq \omega$  where  $\mu = (\mu_1, \dots, \mu_k)$  is a vector of non-negative integers,  $|\mu| = \sum_{j=1}^k \mu_j$ , and  $D^\mu f_u(x) = \partial^{|\mu|} f_u(x) / \partial x_1^{\mu_1} \dots \partial x_k^{\mu_k}$  is the mixed partial derivative of order  $|\mu|$ . The kernel  $K(\cdot)$  satisfies i)  $\int K(x) dx = 1$ ,  $\int x^\mu K(x) dx = 0$  for all  $1 \leq |\mu| \leq \omega - 1$ ,  $\int |x^\mu K(x)| dx < \infty$  for all  $\mu$  with  $|\mu| \leq \omega$ ,  $K(x) \rightarrow 0$  as  $\|x\| \rightarrow \infty$  and  $\sup_{x \in \mathbb{R}^k} \max(1, \|x\|) |D^{e_i} K(x)| < \infty$  for all  $i \leq k$  and  $e_i$  is the  $i$ -th elementary vector in  $\mathbb{R}^k$ . ii)  $K(x)$  is absolutely integrable and has Fourier transform  $\mathfrak{K}(r) = (2\pi)^k \int \exp(ir'x) K(x) dx$  that satisfies  $\int (1 + \|r\|) \sup_{b \geq 1} |\mathfrak{K}(br)| dr < \infty$  where  $i = \sqrt{-1}$ .

Our main results are stated next. All proofs are available in the Auxiliary Appendix published online.

**Theorem 2** *Assume Conditions 1-9 are satisfied. Fix  $x < \infty$  arbitrary and define*

$$\Upsilon_x = \left\{ v \in [-\infty, \infty]^k \mid v = \pi_x v \right\}.$$

*Then, for  $T_n$  defined in (15),  $\sup_{v \in \Upsilon_x} \left| T_n \hat{V}_n(v) - W(v) \right| = o_p(1)$ .*

**Theorem 3** *Assume Conditions 1-10 are satisfied. Fix  $x < 1$  arbitrary and define*

$$\Upsilon_{[0,1]} = \{ w \in \Upsilon_\varepsilon \mid w = \pi_x w \}$$

*where  $\Upsilon_\varepsilon$  is a compact subset of the interior of  $[0, 1]^k$  with volume  $1 - \varepsilon$  for some  $\varepsilon > 0$ . Then,*

$$\sup_{w \in \Upsilon_{[0,1]}} \left| \hat{B}_{\hat{w},n}(w) - B_w(w) \right| = o_p(1).$$

**Theorem 4** *Assume Conditions 1-10 are satisfied. For  $\hat{B}_{\hat{w},n}^*(w)$  defined in (23) it follows that  $\hat{B}_{\hat{w},n}^*(w)$  converges on  $\Upsilon_{[0,1]}$ , defined as in Theorem 3 to a Gaussian process  $B_w(w)$ .*

## D Model Definitions

The model names below summarize variation in *control sets* across propensity score specifications.. All models fit ordered Probit specifications to the change in the discretized intended federal funds rate ( $dDff_t$ ).

- Models a-f; Romer.specificaitons
  - **(Romer baseline)** Baseline specification (a) uses the covariates included in Romer and Romer’s (2004) equation (1), with two modifications: We use the change in the lagged intended federal funds rate instead of the lagged level of the intended federal funds rate; we use the innovation in the unemployment rate, defined as the Greenbook forecast for the unemployment rate in the current quarter minus the unemployment rate in the previous month, instead of the unemployment level used by the Romers. These modifications are meant to eliminate possibly nonstationary regressors. The complete conditioning list includes: the lagged change in the intended federal funds rate, plus the covariates  $graym_t$ ,  $gray0_t$ ,  $gray1_t$ ,  $gray2_t$ ,  $igrym_t$ ,  $igry0_t$ ,  $igry1_t$ ,  $igry2_t$ ,  $gradm_t$ ,  $grad0_t$ ,  $grad1_t$ ,  $grad2_t$ ,  $igrdm_t$ ,  $igrd0_t$ ,  $igrd1_t$ ,  $igrd2_t$ , and our constructed unemployment innovation. For variable names, see Appendix E.
  - **(Restricted Romer)** Specification (b) modifies our baseline specification by eliminating variables with very low significance levels in the multinomial Probit model for the intended rate change. Specifically, we dropped variables with low significance subject to the restriction that if a first-differenced variable from the Romers’ list is retained, then the undifferenced version should appear as well. The retained variable list includes the lagged intended rate change,

gray0<sub>t</sub>, gray1<sub>t</sub>, gray2<sub>t</sub>, igry0<sub>t</sub>, igry1<sub>t</sub>, igry2<sub>t</sub>, grad2<sub>t</sub>, and our constructed unemployment innovation.

- **(Romer lag-quadratic)** Specification (c) adds a quadratic term in the lagged intended federal funds rate change to the restricted model (b).
- **(Romer-discrete baseline/restricted/quadratic)** Specifications (d)-(f) are versions of (a)-(c) which use a discretized variable for the lagged change in the intended federal funds rate.
- Models g - j; Taylor specifications
  - **(Taylor baseline)** Specification (g) uses two lags of  $dff_t$ , 9 lags of the growth rate of real GDP as well as 9 lags of the monthly inflation rate as covariates.
  - **(Taylor lag quadratic)** Specification (h) replaces  $dff_{t-2}$  with  $(dff_{t-1})^2$  in specification (g).
  - **(Taylor-discrete baseline/lag quadratic)** Specifications (i) and (j) are versions of (g) and (h) where covariates based on  $dff_t$  are replaced by covariates based on  $dDff_t$ .



## E Variable Names

$dff_t$	Change in the intended federal funds rate
$Dff_t$	Discretized intended federal funds rate
$dDff_t$	Change in the discretized intended federal funds rate
$innovation_t$	Unemployment innovation
$dDffU_t$	a dummy indicating increases in the intended federal funds rate
$dDffD_t$	a dummy indicating decreases in the intended federal funds rate
$gdp_{t-k}$	$k$ th lag of GDP growth
$inf_{t-k}$	$k$ th lag of inflation
$dIP_t$	change of log of non-seasonally adjusted index of industrial production
$cIP_{t+k}$	$\sum_{j=1}^k dIP_{t+j}$ ; cumulative change in $dIP_t$ .

*From Romer and Romer (2004)*

$graym_t$	Greenbook forecast of the percentage change in real GDP/GNP (at an annual rate) for the previous quarter.
$gray0_t$	Same as above, for current quarter.
$gray1_t$	Same as above, for one quarter ahead.
$gray2_t$	Same as above, for two quarters ahead.

- $igrym_t$  The innovation in the Greenbook forecast for the percentage change in GDP/GNP (at an annual rate) for the previous quarter from the meeting before. The horizon of the forecast for the meeting before is adjusted so that the forecasts for the two meetings always refer to the same quarter.
- $igry0_t$  Same as above, for current quarter.
- $igry1_t$  Same as above, for one quarter ahead.
- $igry2_t$  Same as above, for two quarters ahead.
- $gradm_t$  Greenbook forecast of the percentage change in the GDP/GNP deflator (at an annual rate) for the previous quarter.
- $grad0_t$  Same as above, for current quarter.
- $grad1_t$  Same as above, for one quarter ahead.
- $grad2_t$  Same as above, for two quarters ahead.
- $igrdm_t$  The innovation in the Greenbook forecast for the percentage change in the GDP/GNP deflator (at an annual rate) for the previous quarter from the meeting before. The horizon of the forecast for the meeting before is adjusted so that the forecasts for the two meetings always refer to the same quarter.
- $igrd0_t$  Same as above, for current quarter.
- $igrd1_t$  Same as above, for one quarter ahead.
- $igrd2_t$  Same as above, for two quarters ahead.

## References

- ABADIE, A. (2002): “Bootstrap Tests for Distributional Treatment Effects in Instrumental Variables Models,” *Journal of the American Statistical Association*, pp. 284–292.
- (2005): “Semiparametric Difference-in-Differences Estimators,” *Review of Economic Studies*, 72(1), 1–19.
- ABADIE, A. AND G. W. IMBENS (2009): “Matching on the Estimated Propensity Score,” *mimeo*.
- 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 (2008): “Causal Effects of Monetary Policy Shocks: Semiparametric Conditional Independence Tests with a Multinomial Propensity Score,” *IZA Discussion Paper* 3606.
- ANGRIST, J. D. AND J. PISCHKE (2009): *Mostly Harmless Econometrics: An Empiricist’s Companion* Princeton University Press, Princeton.
- BAI, J. (2003): “Testing Parametric Conditional Distributions of Dynamic Models,” *The Review of Economics and Statistics*, pp. 531–549.
- BERNANKE, B. S., AND A. S. BLINDER (1992): “The Federal Funds Rate and the Channels of Monetary Transmission,” *The American Economic Review*, 82, 901–921.
- BERNANKE, B. S., AND J. BOIVIN (2003): “Monetary Policy in a Data-Rich Environment,” *Journal of Monetary Economics*, 50, 525–546.

- BERNANKE, B. S., J. BOIVIN, AND P. ELIASZ (2005): “Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach,” *The Quarterly Journal of Economics*, pp. 387–422.
- BERNANKE, B. S., AND I. MIHOV (1998): “Measuring Monetary Policy,” *The Quarterly Journal of Economics*, 113, 869–902.
- BIERENS, H. J. (1982): “Consistent Model Specification Tests,” *Journal of Econometrics*, 20, 105–134.
- (1987): “Kernel Estimators of Regression Functions,” in *Advances in Econometrics: Fifth World Congress*, ed. by T. Bewley, pp. 99–144. Cambridge University Press, New York.
- (1990): “A consistent conditional moment test of functional form.,” *Econometrica*, 58, 1443–1458.
- BIERENS, H. J., AND W. PLOBERGER (1997): “Asymptotic theory of integrated conditional moment tests,” *Econometrica*, 65, 1129–1152.
- BLANCHARD, O. J., AND C. M. KAHN (1980): “The Solution of Linear Difference Models under Rational Expectations,” *Econometrica*, 48, 1305–1311.
- CHAMBERLAIN, G. (1982): “The General Equivalence of Granger and Sims Causality,” *Econometrica*, pp. 569–581.
- CHEN, X., AND Y. FAN (1999): “Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series,” *Journal of Econometrics*, 91, 373–401.

- CHRISTIANO, L., 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. 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.
- CLARIDA, R., J. GALI, 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): “Shocks,” *Carnegie-Rochester Conference Series on Public Policy*, 41, 295–364.
- 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.
- DELGADO, M. A., AND W. STUTE (2008): “Distribution Free Specification Tests of Conditional Models,” *Journal of Econometrics*, 143, 37–55.
- DUFOUR, J.-M., AND E. RENAULT (1998): “Short Run and Long Run Causality in Time Series: Theory,” *Econometrica*, pp. 1099–1125.
- DUFOUR, J.-M., AND D. TESSIER (1993): “On the relationship between impulse response analysis, innovation accounting and Granger causality,” *Economics Letters*, 42, 327–333.
- EICHENGREEN, B., M. W. WATSON, AND R. S. GROSSMAN (1985): “Bank Rate Policy Under the Interwar Gold Standard: A Dynamic Probit Model,” *The Economic Journal*, 95, 725–745.

- FLORENS, J.-P., AND M. MOUCHART (1982): “A Note on Non-Causality,” *Econometrica*, pp. 582–591.
- FLORENS, J.-P., AND M. MOUCHART (1985): “A Linear Theory for Noncausality,” *Econometrica*, pp. 157–176.
- GENBERG, H., AND S. GERLACH (2004): “Estimating Central Bank Reaction Function with Ordered Probit: A Note,” *mimeo*.
- GORDON, D. B., AND E. M. LEEPER (1994): “The Dynamic Impacts of Monetary Policy: An Exercise in Tentative Identification,” *The Journal of Political Economy*, 102, 1228–1247.
- HAHN, J. (1999): “How informative is the initial condition in the dynamic panel model with fixed effects,” *Journal of Econometrics*, 93, 309–326.
- HAMILTON, J. D., AND O. JORDA (2002): “A Model of the Federal Funds Rate Target,” *Journal of Political Economy*, pp. 1135–1167.
- HANSEN, B. E. (1996): “Inference when a nuisance parameter is not identified under the null hypothesis,” *Econometrica*, 64, 413–430.
- HECKMAN, J. J., H. ICHIMURA, AND P. E. TODD (1997): “Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme,” *Review of Economic Studies*, 64, 605–654.
- 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., AND G. IMBENS (2004): “The Propensity Score with Continuous Treatments,” *Berkeley Department of Economics, mimeo*.
- HIRANO, K., G. IMBENS, AND G. RIDDER (2003): “Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score,” *Econometrica*, 71, 1161–1189.
- HU, L., AND P. C. PHILLIPS (2002a): “Dynamics of the Federal Funds Target Rate: A Nonstationary Discrete Choice Approach,” *Cowles Foundation Discussion Paper No. 1365*.
- (2002b): “Nonstationary Discrete Choice,” *Cowles Foundation Discussion Paper No 1364*.
- IMBENS, G. W. (2000): “The Role of the Propensity Score in Estimating Dose-Response Functions,” *Biometrika*, 87, 706–710.
- KHMALADZE, E. (1981): “Martingale Approach in the Theory of Goodness-of-fit Tests,” *Theory of Probability and Its Applications*, pp. 240–257.
- (1988): “An Innovation Approach to Goodness-of-Fit Tests in  $R^m$ ,” *Annals of Statistics*, pp. 1503–1516.
- KHMALADZE, E. V. (1993): “Goodness of Fit Problem and Scanning Innovation Martingales,” *The Annals of Statistics*, pp. 789–829.
- KOENKER, R., AND Z. XIAO (2002): “Inference of the Quantile Regression Process,” *Econometrica*, 70, 1583–1612.
- KOUL, H. L., AND W. STUTE (1999): “Nonparametric Model Checks for Time Series,” *Annals of Statistics*, pp. 204–236.

- LECHNER, M. (2000): “Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption,” in *Economic Evaluations of Active Labour Market Policies in Europe*, ed. by M. Lechner, and F. Pfeiffer. Physica, Heidelberg.
- LECHNER, M. (2004): “Sequential Matching Estimation of Dynamic Causal Models,” *Discussion Paper 2004-06, Department of Economics, University of St. Gallen*.
- LEEPER, E. M. (1997): “Narrative and VAR approaches to Monetary Policy: Common Identification Problems,” *Journal of Monetary Economics*, pp. 641–657.
- LEEPER, E. M., C. A. SIMS, AND T. ZHA (1996): “What Does Monetary Policy Do?,” *Brookings Papers on Economic Activity*, 1996, 1–78.
- LINTON, O., AND P. GOZALO (1999): “Conditional Independence Restrictions: Testing and Estimation,” *mimeo*.
- LOK, J., R. GILL, A. V. DER VAART, AND J. ROBINS (2004): “Estimating the causal effect of a time-varying treatment on time-to-event using structural nested failure time models,” *Statistica Neerlandica*, 58, 271–295.
- LUBIK, T. A., AND F. SCHORFHEIDE (2003): “Computing sunspot equilibria in linear rational expectations models,” *Journal of Economic Dynamics and Control*, 28, 273–285.
- (2004): “Testing for Indeterminacy: An Application to U.S. Monetary Policy,” *The American Economic Review*, 94, 190–217.



- MAMMEN, E. (1993): “Bootstrap and the Wild Bootstrap for High Dimensional Linear Models,” *Annals of Statistics*, 21, 255–285.
- MOENCH, E., AND H. UHLIG (2005): “Towards a Monthly Business Cycle Chronology for the Euro Area,” *Journal of Business Cycle Measurement and Analysis*, 2, 43–69.
- PARK, J. Y., AND P. C. PHILLIPS (2000): “Nonstationary Binary Choice,” *Econometrica*, pp. 1249–1280.
- POLLARD, D. (1984): *Convergence of Stochastic Processes*. Springer-Verlag New York, Inc.
- ROBINS, J. M., S. GREENLAND, AND F.-C. HU (1999): “Estimation of the Causal Effect of a Time-Varying Exposure on the Marginal Mean of a Repeated Binary Outcome,” *Journal of the American Statistical Association*, pp. 687–712.
- ROBINS, J. M., S. D. MARK, AND W.K. NEWEY (1992): “Estimating Exposure Effects by Modelling the Expectation of Exposure Conditional on Confounders,” *Biometrics*, 48, pp. 479–495.
- ROMER, C. D., AND D. H. ROMER (2004): “A New Measure of Monetary Shocks: Derivation and Implications,” *The American Economic Review*, 94, 1055–1084.
- ROSENBAUM, P. R., AND D. B. RUBIN (1983): “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika*, pp. 41–55.
- ROSENBLATT, M. (1952): “Remarks on a Multivariate Transform,” *The Annals of Mathematical Statistics*, 23(3), 470–472.
- ROTEMBERG, J. J., AND M. WOODFORD (1997): “An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy,” *NBER Macroeconomic Annual*, 12, 297–346.

- RUBIN, D. B. (1974): “Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies,” *Journal of Educational Psychology*, pp. 688–701.
- (1978): “Bayesian Inference for Causal Effects: The Role of Randomization,” *The Annals of Statistics*, 6, 34–58.
- SHAPIRO, M. D. (1994): “Federal Reserve Policy: Cause and Effect,” in *Monetary Policy*, ed. by G. N. Mankiew, pp. 307–334. University of Chicago Press.
- SIMS, C., AND T. ZHA (2006): “Does Monetary Policy Generate Recessions,” *Macroeconomic Dynamics*, 10, 231–272.
- SIMS, C. A. (1972): “Money, Income and Causality,” *American Economic Review*, pp. 540–562.
- (1980): “Macroeconomics and Reality,” *Econometrica*, pp. 1–48.
- (2001): “Solving Linear Rational Expectations Models,” *Computational Economics*, 20, 1–20.
- SONG, K (2009): “Testing Conditional Independence via Rosenblatt Transforms,” *Annals of Statistics*, 37, 4011–4045.
- STOCK, J. H., AND M. W. WATSON (2002a): “Forecasting Using Principle Components from a Large Number of Predictors,” *Journal of the American Statistical Association*, 97, 1167–1179.
- (2002b): “Macroeconomic Forecasting Using Diffusion Indexes,” *Journal of Business and Economic Statistics*, 20, 147–162.
- STRONGIN, S. (1995): “The Identification of Monetary Policy Disturbances Explaining the Liquidity Puzzle,” *Journal of Monetary Economics*, 35, 463–497.

SU, L., AND H. WHITE (2003): “Testing Conditional Independence via Empirical Likelihood,” *UCSD Discussion Paper 2003-14*.

STUTE, W., S. THIES, AND L.-X. ZHU (1998): “Model Checks for Regression: An Innovation Process Approach,” *Annals of Statistics*, pp. 1916–1934.

TAYLOR, J. B. (1993): “Discretion versus policy rules in practice,” *Carnegie-Rochester Conference Series on Public Policy*, 39, 195–214.

WHITE, H. (2006): “Time Series Estimation of the Effects Natural Experiments,” *Journal of Econometrics*, 135, 527–566.

WHITE, H. (2009): “Retrospective Estimation of Causal Effects through Time,” *mimeo*.

Table 1a: Parametric Sims-causality Tests for models using lagged  $dfi_t$

Lead	Model				
	(a)	(b)	(c)	(g)	(h)
1	1.08	0.99	1.50	1.64	2.34 **
2	0.25	0.17	0.73	0.60	1.45
3	-0.40	-0.48	0.04	-0.37	0.36
4	-1.30	-1.55	-0.40	-1.40	-0.15
5	-0.93	-1.16	-0.23	-1.32	-0.32
6	-1.42	-1.69 *	-0.89	-2.06 **	-1.27
7	-2.21 **	-2.45 **	-1.66 *	-2.69 ***	-1.97 **
8	-3.67 ***	-3.84 ***	-3.19 ***	-4.16 ***	-3.45 ***
9	-3.92 ***	-4.01 ***	-3.36 ***	-4.72 ***	-3.97 ***
10	-3.86 ***	-3.98 ***	-3.41 ***	-4.82 ***	-4.20 ***
11	-4.03 ***	-4.12 ***	-3.66 ***	-4.82 ***	-4.34 ***
12	-4.02 ***	-4.03 ***	-3.90 ***	-4.93 ***	-4.70 ***

Notes: The table reports t-statistics for parametric Sims-causality tests for the response of

the change in the log of the non-seasonally-adjusted index of industrial production to

monetary policy shocks. Columns report results using alternative models for the policy

propensity score. Model details are summarized in the model definitions appendix.

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Table 1b: Parametric Sims-causality Tests for models using lagged  $dDff_t$

Lead	Model				
	(d)	(e)	(f)	(i)	(j)
1	1.18	1.04	0.98 *	1.88	1.92 *
2	0.94	0.96	0.92	1.41	1.50
3	-0.37	-0.30	-0.40	0.05	0.14
4	-0.94	-0.92	-0.99	-0.49	-0.42
5	-0.49	-0.52	-0.60	-0.34	-0.27
6	-0.62	-0.63	-0.71	-0.85	-0.72
7	-1.59	-1.55 *	-1.65 *	-1.65 *	-1.49
8	-2.78 ***	-2.70 ***	-2.78 ***	-2.75 ***	-2.55 **
9	-3.02 ***	-2.97 ***	-3.05 ***	-3.29 ***	-3.05 ***
10	-2.83 ***	-2.81 ***	-2.83 ***	-3.32 ***	-3.02 ***
11	-3.28 ***	-3.23 ***	-3.25 ***	-3.53 ***	-3.27 ***
12	-3.37 ***	-3.26 ***	-3.27 ***	-3.62 ***	-3.45 ***

Notes: The table reports t-statistics for parametric Sims-causality tests for the response of the

change in the log of the non-seasonally-adjusted index of industrial production to monetary

policy shocks. Columns report results using alternative models for the policy propensity

score. Model details are summarized in the model definitions appendix.

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Table 2a: Specification Tests for models using lagged  $dff_t$ 

Variable	Model									
	(a)		(b)		(c)		(g)		(h)	
	p-val	Sig	p-val	Sig	p-val	Sig	p-val	Sig	p-val	Sig
$dff_t$	0.235		0.167		0.557		0.000	***	0.120	
$graym_t$	0.655		0.346		0.880		0.844		0.578	
$gray0_t$	0.118		0.696		0.515		0.660		0.090	*
$gray1_t$	0.522		0.666		0.867		0.824		0.305	
$gray2_t$	0.203		0.437		0.856		0.975		0.645	
$igrym_t$	0.509		0.703		0.439		0.631		0.727	
$igr0_t$	0.609		0.908		0.915		0.004	***	0.001	***
$igr1_t$	0.231		0.627		0.868		0.117		0.100	
$igr2_t$	0.209		0.472		0.621		0.033	**	0.006	***
$gradm_t$	0.626		0.574		0.472		0.143		0.673	
$grad0_t$	0.176		0.721		0.882		0.297		0.631	
$grad1_t$	0.505		0.394		0.060	*	0.478		0.349	
$grad2_t$	0.362		0.431		0.111		0.268		0.432	
$igrdm_t$	0.496		0.713		0.397		0.615		0.705	
$igrd0_t$	0.789		0.652		0.704		0.836		0.645	
$igrd1_t$	0.185		0.299		0.510		0.536		0.594	
$igrd2_t$	0.089	*	0.265		0.248		0.087	*	0.132	
$innovation_t$	0.535		0.043	**	0.687		0.451		0.581	
$gdp_{t-1}$	0.167		0.134		0.272		0.385		0.450	
$gdp_{t-2}$	0.715		0.219		0.302		0.614		0.308	
$gdp_{t-3}$	0.950		0.295		0.653		0.135		0.800	
$gdp_{t-4}$	0.592		0.644		0.235		0.922		0.791	
$gdp_{t-5}$	0.060	*	0.320		0.508		0.613		0.539	
$gdp_{t-6}$	0.538		0.760		0.386		0.540		0.163	
$gdp_{t-7}$	0.738		0.588		0.371		0.646		0.820	
$gdp_{t-8}$	0.737		0.663		0.872		0.604		0.228	
$gdp_{t-9}$	0.311		0.306		0.358		0.535		0.656	
$Inf_{t-1}$	0.744		0.760		0.649		0.609		0.800	
$Inf_{t-2}$	0.823		0.829		0.802		0.699		0.929	
$Inf_{t-3}$	0.866		0.571		0.337		0.495		0.338	
$Inf_{t-4}$	0.318		0.262		0.094	*	0.585		0.431	
$Inf_{t-5}$	0.407		0.731		0.704		0.215		0.547	
$Inf_{t-6}$	0.128		0.460		0.327		0.136		0.772	
$Inf_{t-7}$	0.721		0.290		0.682		0.429		0.985	
$Inf_{t-8}$	0.093	*	0.273		0.187		0.118		0.652	
$Inf_{t-9}$	0.736		0.502		0.617		0.394		0.897	

Notes: The table reports p-values for the semiparametric causality tests VM defined in (17) and based

on the moment condition (8) with  $\phi(z_{it}, v_2)$  equal to  $1\{z_{it} \leq v_2\}$ . Each line uses the specified variable as

$z_{it}$ . Variables are defined in the variable names appendix. Columns report results using alternative models

for the policy propensity score. Model details are summarized in the model definitions appendix. P-values

use a bootstrap of the transformed test statistic. See text for details.

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Table 2b: Specification Tests for models using lagged  $dDff_t$ 

Variable	Model									
	(d)		(e)		(f)		(i)		(j)	
	p-val	Sig	p-val	Sig	p-val	Sig	p-val	Sig	p-val	Sig
$dDff_t$	0.028	**	0.133		0.000	***	0.197		0.013	**
$graym_t$	0.762		0.547		0.750		0.848		0.850	
$gray0_t$	0.031	**	0.114		0.230		0.135		0.135	
$gray1_t$	0.285		0.548		0.974		0.358		0.343	
$gray2_t$	0.183		0.664		0.622		0.529		0.502	
$igrym_t$	0.539		0.138		0.502		0.287		0.371	
$igry0_t$	0.580		0.454		0.745		0.010	**	0.009	***
$igry1_t$	0.212		0.570		0.686		0.035	**	0.033	**
$igry2_t$	0.278		0.124		0.422		0.007	***	0.008	***
$gradm_t$	0.738		0.145		0.613		0.806		0.851	
$grad0_t$	0.295		0.563		0.272		0.703		0.834	
$grad1_t$	0.682		0.151		0.636		0.381		0.345	
$grad2_t$	0.412		0.172		0.220		0.371		0.361	
$igrdm_t$	0.507		0.136		0.492		0.316		0.371	
$igrd0_t$	0.651		0.600		0.352		0.785		0.788	
$igrd1_t$	0.220		0.395		0.558		0.358		0.408	
$igrd2_t$	0.034	**	0.162		0.081	*	0.091	*	0.093	*
$innovation_t$	0.642		0.190		0.492		0.654		0.747	
$gdp_{t-1}$	0.143		0.198		0.174		0.439		0.511	
$gdp_{t-2}$	0.740		0.252		0.246		0.720		0.676	
$gdp_{t-3}$	0.967		0.909		0.697		0.410		0.332	
$gdp_{t-4}$	0.526		0.599		0.532		0.805		0.858	
$gdp_{t-5}$	0.071	*	0.460		0.228		0.601		0.714	
$gdp_{t-6}$	0.556		0.403		0.757		0.741		0.683	
$gdp_{t-7}$	0.787		0.168		0.623		0.812		0.821	
$gdp_{t-8}$	0.823		0.815		0.852		0.264		0.285	
$gdp_{t-9}$	0.197		0.304		0.144		0.647		0.660	
$Inf_{t-1}$	0.533		0.607		0.705		0.838		0.832	
$Inf_{t-2}$	0.638		0.840		0.582		0.619		0.793	
$Inf_{t-3}$	0.861		0.566		0.740		0.857		0.801	
$Inf_{t-4}$	0.253		0.096	*	0.148		0.769		0.768	
$Inf_{t-5}$	0.204		0.882		0.310		0.151		0.203	
$Inf_{t-6}$	0.150		0.318		0.279		0.198		0.340	
$Inf_{t-7}$	0.718		0.464		0.417		0.737		0.903	
$Inf_{t-8}$	0.068	*	0.415		0.138		0.278		0.325	
$Inf_{t-9}$	0.755		0.686		0.775		0.648		0.778	

Notes: The table reports p-values for the semiparametric causality tests VM defined in (17) and based

on the moment condition (8) with  $\phi(z_{it}, v_2)$  equal to  $1\{z_{it} \leq v_2\}$ . Each line uses the specified variable as

$z_{it}$ . Variables are defined in the variable names appendix. Columns report results using alternative models

for the policy propensity score. Model details are summarized in the model definitions appendix. P-values

use a bootstrap of the transformed test statistic. See text for details.

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Table 3a: Semiparametric Causality Tests using lagged  $dff_t$ , for Up and Down Policy Changes

Lead	Model									
	(a)		(b)		(c)		(g)		(h)	
	p-val	Sig	p-val	Sig	p-val	Sig	p-val	Sig	p-val	Sig
1	0.616		0.748		0.398		0.902		0.850	
2	0.731		0.888		0.611		0.544		0.780	
3	0.325		0.500		0.500		0.853		0.533	
4	0.109		0.247		0.275		0.462		0.200	
5	0.141		0.450		0.548		0.560		0.216	
6	0.127		0.347		0.726		0.428		0.125	
7	0.040	**	0.158		0.697		0.205		0.094	*
8	0.053	*	0.116		0.144		0.094	*	0.005	***
9	0.124		0.157		0.063	*	0.044	**	0.010	**
10	0.092	*	0.048	**	0.020	**	0.038	**	0.006	***
11	0.025	**	0.008	***	0.019	**	0.018	**	0.002	***
12	0.062	*	0.042	**	0.020	**	0.021	**	0.001	***

Notes: The table reports p-values for the semiparametric causality tests VM defined in (17) and based on

the moment condition (7) with  $\phi(U_t, v)$  equal to  $1\{y_t \leq v\}$ . In this implementation,  $D_t$  is a bivariate vector containing dummy variables for an up or down movement of  $dDff_t$ . Columns report results using alternative models for the policy propensity score. Model details are summarized in the model definitions appendix.

P-values use a bootstrap of the transformed test statistic. See text for details.

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%



Table 3b: Semiparametric Causality Tests using lagged  $dDff_t$ , for Up and Down Policy Changes

Lead	Model									
	(d)		(e)		(f)		(i)		(j)	
	p-val	Sig	p-val	Sig	p-val	Sig	p-val	Sig	p-val	Sig
1	0.552		0.555		0.461		0.645		0.613	
2	0.561		0.740		0.674		0.299		0.313	
3	0.150		0.177		0.143		0.371		0.403	
4	0.100		0.108		0.096	*	0.070	*	0.092	*
5	0.100		0.180		0.166		0.255		0.317	
6	0.085	*	0.175		0.158		0.167		0.241	
7	0.054	*	0.117		0.125		0.066	*	0.074	*
8	0.047	**	0.083	*	0.079	*	0.014	**	0.020	**
9	0.114		0.097	*	0.064	*	0.014	**	0.016	**
10	0.167		0.100		0.026	**	0.025	**	0.026	**
11	0.054	*	0.030	**	0.020	**	0.012	**	0.013	**
12	0.065	*	0.032	**	0.013	**	0.005	***	0.008	***

Notes: The table reports p-values for the semiparametric causality tests VM defined in (17) and based on

the moment condition (7) with  $\phi(U_t, v)$  equal to  $1\{y_t \leq v\}$ . In this implementation,  $D_t$  is a bivariate vector containing dummy variables for an up or down movement of  $dDff_t$ . Columns report results using alternative models for the policy propensity score. Model details are summarized in the model definitions appendix.

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Table 4: Effects of a Surprise Decrease in the Federal Funds Target Rate

Lead	Model							
	(a)		(c)		(e)		(h)	
	p-val	Sig	p-val	Sig	p-val	Sig	p-val	Sig
1	0.417		0.210		0.398		0.597	
2	0.740		0.395		0.520		0.775	
3	0.896		0.420		0.654		0.618	
4	0.209		0.212		0.274		0.264	
5	0.508		0.470		0.491		0.681	
6	0.673		0.824		0.527		0.638	
7	0.393		0.675		0.523		0.665	
8	0.315		0.166		0.404		0.398	
9	0.743		0.092	*	0.603		0.115	
10	0.095	*	0.020	**	0.100		0.052	*
11	0.036	**	0.022	**	0.072	*	0.069	*
12	0.176		0.044	**	0.105		0.178	

Notes: The table reports p-values for the semiparametric causality tests VM

defined in (17) and based on the moment condition (7) with  $\phi(U_t, v)$  equal to  $1\{y_t \leq v\}$ . In this implementation,  $D_t$  is a dummy variable defined to equal one when the intended federal funds rate is changed downwards. Columns report results using alternative models for the policy propensity score. Model details are summarized in the model definitions appendix. P-values use a bootstrap of the transformed test statistic. See text for details.

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Table 5: Effects of a Surprise Increase in the Federal Funds Target Rate

Lead	Model							
	(a)		(c)		(e)		(h)	
	p-val	Sig	p-val	Sig	p-val	Sig	p-val	Sig
1	0.582		0.610		0.484		0.874	
2	0.503		0.778		0.681		0.494	
3	0.154		0.407		0.093 *		0.336	
4	0.082 *		0.398		0.066 *		0.169	
5	0.086 *		0.401		0.098 *		0.079 *	
6	0.046 **		0.361		0.078 *		0.052 *	
7	0.020 **		0.466		0.068 *		0.027 **	
8	0.026 **		0.169		0.041 **		0.004 ***	
9	0.043 **		0.167		0.044 **		0.007 ***	
10	0.114		0.349		0.136		0.013 **	
11	0.037 **		0.084 *		0.036 **		0.001 ***	
12	0.027 **		0.040 **		0.041 **		0.000 ***	

Notes: The table reports p-values for the semiparametric causality tests VM

defined in (17) and based on the moment condition (7) with  $\phi(U_t, v)$  equal to  $1\{y_t \leq v\}$ .

In this implementation,  $D_t$  is a dummy variable defined to equal one when

the intended federal funds rate increases. Columns report results using alternative

models for the policy propensity score. Model details are summarized in the

model definitions appendix. P-values use a bootstrap of the transformed test

statistic. See text for details.

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%