IZA

DISCUSSION PAPER SERIES

IZA DP No. 3606

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

Joshua D. Angrist
Guido M. Kuersteiner

July 2008

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

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

Joshua D. Angrist
MIT and IZA

Guido M. Kuersteiner
UC Davis

Discussion Paper No. 3606
July 2008

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 3606
July 2008

ABSTRACT

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

Macroeconomists have long been concerned with the causal effects of monetary policy. When the identification of causal effects is based on a selection-on-observables assumption, non-causality amounts to the conditional independence of outcomes and policy changes. This paper develops a semiparametric test for conditional independence in time series models linking a multinomial policy variable with unobserved potential outcomes. Our approach to conditional independence testing is motivated by earlier parametric tests, as in Romer and Romer (1989, 1994, 2004). The procedure developed here 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. This leads to a generalized definition of Sims (1980) causality. A technical contribution is the development of root-T consistent distribution-free inference methods for full conditional independence testing, appropriate for dependent data and allowing for first-step estimation of the propensity score.

JEL Classification: E52, C22, C31

Keywords: monetary policy, propensity score, multinomial treatments, causality

Corresponding author:

Joshua Angrist
Department of Economics
MIT E52-353
50 Memorial Drive
Cambridge, MA 02139-437
USA
E-mail: angrist@mit.edu

* This paper is a revised version of NBER working paper 10975, December 2004. We thank Christina Romer for sharing the Romers' 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 Columbia University, Harvard-MIT, 2004 NSF/NBER Time Series Conference, Rochester, Rutgers, the Triangle Econometrics Workshop, Wisconsin and the 2nd conference on evaluation research in Mannheim for helpful comments. Kuersteiner gratefully acknowledges support from NSF grants SES-0095132 and SES-0523186.

1 Introduction

The possibility of a 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 surprise variables 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 a compelling exclusion restriction. This assumption is equivalent to postulating independent structural innovations in structural vector autoregressions (SVAR) which have taken center stage in the analysis of monetary policy effects. Recent contributions to this literature include Bernanke and Blinder (1992), Christiano, Eichenbaum and Evens (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. Essentially, 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 complete model of both the processes determining outcomes and the process determining policies towards a model of the process determining policy decisions alone. 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 remaining part of the system unspecified. This limited focus should increase robustness. For example, we do not need to specify the functional form or lag length in a model for GDP growth. Rather, we need be concerned solely with the horizon and variables relevant for Federal Open market Committee (FOMC) decision-making, issues about which there is considerable institutional knowledge. Moreover, the multinomial nature of some policy variables 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 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 estimation 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), and Abadie (2005), 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.[1] 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, suitable for dependent data, and 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.[2] 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

[1] 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.

[2] 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 (2002) and Koenker and Xiao (2002) .

therefore use a Rosenblatt (1952) transformation of the data in addition to the Khmaladze transformation[3]. This combination of methods seems to perform well, at least for the low-dimensional 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[4] of alternative causality concepts and test statistics is illustrated through a re-analysis of the Romer and Romer (2004) data in Section 5. 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.

2 Notation and Framework

Causal effects are defined here using the Rubin (1974) notion of potential outcomes. The potential outcomes concept originated in experimental studies where the investigator has control over the assignment of treatments, but is now widely used in observational studies. Our definition of causality relies on distinguishing the 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 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 any serial correlation or covariates, the causal effect of a treatment or policy action is unambiguously defined as $Y_{1t} - Y_{0t}$. It is clear that this effect can never be measured in practice. Researchers therefore focus on either the average effect $E(Y_{1t} - Y_{0t})$, or the effect in treated periods, $E(Y_{1t} - Y_{0t}|D_t = 1)$. We refer to both of these as the average causal effect of policy action D_t, since under our identifying assumptions they are the same. 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 includes repeated observations on treatment and outcome variables. 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 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

[3] In recent work, independent of ours, Delgado and Stute (2005) discuss a specification test that also combines the Khmaladze and Rosenblatt transforms.

[4] A small Monte Carlo study can be found in our NBER working paper Angrist and Kuersteiner (2004).

are not part of the null hypothesis of no causal effect of D_t. Let $\bar{X}_t = (X_t, ..., X_{t-k}, ...)$ denote the covariate path, with similar definitions for $\bar{Y}_t$ and $\bar{D}_t$. We assume that the information used by policy makers at time t, denoted $\mathcal{F}_t$, is contained in the public record or otherwise available to researchers. 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 $\Pi_t : \bigotimes_{i=1}^{\dim(\chi_t)} \mathbb{R}^\infty \to \mathbb{R}^{k_2}$ of the entire observable history of the joint process. For the purposes of empirical work, the mapping Π_t is 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 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 ε_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 $D_t = \psi(D(z_t, t), \varepsilon_t)$, where ψ is a general mapping. Without loss of generality we can assume that ε_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 ψ takes values in the set of functions Ψ_t. A common specification in the literature on monetary policy is a Taylor (1993) rule for the nominal interest rate. In this literature, ψ 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 ε_t.

A second key assumption is that the stochastic component of the policy function, ε_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 $\psi(D(z_t, t), \varepsilon_t)$. 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 $D_t = \psi(D(z_t, t), \varepsilon_t) = d$, 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 ε_{t+j-1}, that is, random shocks that occur between time t and $t + j$. When we imagine changing d or ψ to generate potential outcomes, the sequence of intervening shocks is held fixed. This is discussed further in Example 1, below. It's also worth noting that our setup focuses on the effect of a single policy shock on subsequent outcomes. 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.

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 occur in period $t+j$. In particular, $Y_{t,j}^{\psi}(d)$ and $Y_{t+1,j-1}^{\psi}(d')$ may differ because $Y_{t,j}^{\psi}(d)$ does not constrain the policy in period $t+1$ to equal d' and $Y_{t+1,j-1}^{\psi}(d')$ does not constrain the policy in period t to equal d, a point that will be further illustrated in Example 1 below.

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, ψ, t, j.

In the simple situation studied by Rubin (1974), the no-effects null hypothesis states that $Y_{0t} = Y_{1t}$.[5] Our approach to causality testing leaves $Y_{t,j}^{\psi}(d)$ unspecified. In contrast, it's common practice in econometrics to model the joint distribution of the vector of outcomes and policy variables (χ_t) as a function of lagged and exogenous variables or innovations in variables, and so it's worth thinking about what potential outcomes would be in this case. When economic theory provides a model for χ_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 χ_t as a function of past structural innovations $\bar{\varepsilon}_t = (\varepsilon_t, \varepsilon_{t-1}, ...)$ in the policy function and structural innovations $\bar{\eta}_t = \left(\eta_t, \eta_{t-1}, ...\right)$ in the rest of the economy. Further assuming that $\psi(D(z_t,t), \varepsilon_t) = d$ can be solved for ε_t such that for some function ψ^*, $\varepsilon_t = \psi^*(D(z_t,t), d)$ we can then partition $\tilde{\chi}_t = \left(\tilde{Y}_t, \tilde{X}_t, \tilde{D}_t\right)$ 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 $Y_{t,j}^{\psi}(d) = \tilde{Y}_{t+j}\left(\varepsilon_{t+j}, ...\varepsilon_{t+1}, \psi^*(\tilde{D}_t, d), \bar{\varepsilon}_{t-1}, \bar{\eta}_t\right)$[6]. 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 $D(.,.)$ and on the distribution of ε_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

[5] 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 $D_t = D_{t-1} = ... = 0$. They 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.

[6] When $D_t = D(z_t,t) + \varepsilon_t$, $\psi^*(D(z_t,t), d) = d - D(z_t,t)$. However, the function ψ^* may not always exist. Then, it may be more convenient to index potential outcomes directly as functions of ε_t rather than d. In that case, one could define $Y_{t,j}^{\psi}(e) = \tilde{Y}_{t+j}(\varepsilon_{t+j}, ...\varepsilon_{t+1}, e, \bar{\varepsilon}_{t-1}, \bar{\eta}_t)$ 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 $Y_{t,j}^{\psi}(d)$.

an obvious way. As recently highlighted by Clarida, Gali and Gertler (2000) and Lubik and Schorfheide (2003, 2004), however, New Keynesian monetary models have multiple equilibria under certain interest rate targeting rules. Lubik and Schorfheide (2003) in particular show that the effect of policy shocks ε_t may not be unique for some models and parameter combinations. In this case, potential outcomes are also non-unique, but we can accommodate non-uniqueness by allowing $Y_{t,j}^{\psi}(d)$ to be a set-valued random variable. Lubik and Schorfheide (2003) provide an algorithm that can be used to compute potential outcomes for linear rational expectations models even when there are multiple equilibria. Multiplicity of equilibria is compatible with Condition 1 as long as the multiplicity disappears under the Null hypothesis of no causal effect. Moreover, uniqueness of equilibria under the no-effects null only needs to hold for the component $\tilde{Y}_t(\bar{\varepsilon}_t, \bar{\eta}_t)$ of $\tilde{\chi}_t = \left(\tilde{Y}_t, \tilde{X}_t, \tilde{D}_t\right)$. In the context of testing for the effects of monetary policy on the real economy, this means that under the null hypothesis only real variables need to be uniquely determined, while nominal variables such as the inflation rate still may be subject to multiplicity.

The link between the potential outcomes concept and structural macroeconomic models can be made more specific using Bernanke and Blinder's (1992) SVAR model for the effects of money (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 *Suppose that the SVAR takes the form* $\Gamma_0 \chi_t = -\Gamma(L)\chi_t + (\eta_t', \varepsilon_t)'$ *where* Γ_0 *is a matrix of constants conformable to* χ_t *and* $\Gamma(L) = \Gamma_1 L + ... + \Gamma_p L^p$ *is a lag polynomial such that* $C(L) := (\Gamma_0 + \Gamma(L))^{-1} = \sum_{k=0}^{\infty} C_k L^k$ *exists. The policy innovations are denoted by* ε_t *and other structural innovations are* η_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*,* ε_t *and* η_t*. In this setup, potential outcomes are defined as*

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

Potential outcomes answer the following question: assume that everything else equal, which in this case means keeping ε_{t+j-k} *and* η_{t+j-k} *fixed for* $k \neq j$*, how would the outcome variable* Y_{t+j} *change if we change the policy innovation from* ε_t *to* d*? The sharp null hypothesis of no causal effect holds if and only if* $c_{y\varepsilon,j} = 0$ *for all* j*. This is the familiar restriction that the impulse response function be identically equal to zero. In general,* $Y_{t+1,j-1}^{\psi}(d') = \sum_{k=0,k\neq j-1}^{\infty} c_{y\varepsilon,k}\varepsilon_{t+j-k} + \sum_{k=0}^{\infty} c_{y\eta,k}\eta_{t+j-k} + c_{y\varepsilon,j-1}d'$ *differs from* $Y_{t,j}^{\psi}(d)$ *,except when the potential outcomes are evaluated at the realized policy innovations* ε_t *and* ε_{t+1}*, in which case* $Y_{t,j}^{\psi}(\varepsilon_t) = Y_{t+1,j-1}^{\psi}(\varepsilon_{t+1}) = Y_t$ *or under the null hypothesis of no causal effects, where* $Y_{t+1,j-1}^{\psi}(d') = Y_{t,j}^{\psi}(d) = Y_{t+j}$.

Definition 1 extends the conventional potential outcome framework in a number of important ways. A 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 by statistically conditioning on the history of observed policies, covariates and outcomes, so that in practice when we discuss potential outcomes, we have in mind 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, ε_t. The dependence of potential outcomes on the distribution of ε_t is captured by ψ. Finally, the fact that potential outcomes depend on ψ 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:*

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

The selection on observable assumption says that policies are independent of potential outcomes after appropriate conditioning. Note also that Condition 2 implies that $Y_{t,1}^{\psi}(d), Y_{t,2}^{\psi}(d), ... \perp \varepsilon_t | z_t$. 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 ε_t. The variation in ε_t is 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, and 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 finally 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 $Y_{t,j}^{\psi\prime}(d') = Y_{t,j}^{\psi}(d) = Y_{t+j}$. Using this to substitute in Condition 2 produces the key testable conditional independence assumption, written in terms of observable distributions as:

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

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

Identification Condition 2 plays a central role in the applied literature on testing the effects of monetary policy. Of course, Condition 2 is a strong restriction. Nevertheless, a corresponding version of it has been widely used in SVAR analysis of monetary policy shocks. Bernanke and Blinder (1992), Gordon and Leeper (1994), Christiano, Eichenbaum and Evans (1996, 1999), Bernanke and Mihov (1998) all 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 Γ_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 ε_t and η_t are independent this implies that Condition 2 holds in Example 1. To see this, note that conditional on z_t the distribution of D_t only depends on ε_t whose distribution in turn is independent of z_t and thus all past and future ε_t and all η_t. Christiano, Eichenbaum and Evans (1999) discuss a variety of different specifications within the SVAR literature that are based on recursive identification. The key assumption in all these approaches is that an instantaneous response of the conditioning variables z_t to the policy shock ε_t can be ruled out a priori. Leeper, Sims and Zha (1996) and Sims and Zha (2006) argue that these assumptions are often not satisfied and propose identification based on restrictions that involve the entire system matrix Γ_0. When simultaneity is indeed a problem, in other words when the distribution of ε_t conditional on z_t does depend on z_t then Condition 2 can not hold in general because potential outcomes $Y_{t,j}^{\psi}(d)$ generally are not independent of z_t.

Tests based on Condition 1 can be seen as testing a generalized version of Sims causality similar to the one 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 $Y_{t+1}, ..., Y_{t+k}, ... \perp D_t | \bar{Y}_t, \bar{D}_{t-1}$. Chamberlain (1982) and Florens and Mouchart (1982, 1985) show that under plausible regularity conditions this is equivalent to generalized Granger non-causality, i.e.,

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

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

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

This result was shown for the case of linear processes by Dufour and Tessier (1993), but seems to have received little attention in the literature.[7] We summarize the non-equivalence of Sims and Granger causality in the following theorem:

Theorem 1 *Let* χ_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.

A scenario with Granger non-causality but Sims causality is of potential relevance in the debate over money-output causality. Suppose y_t is output, x_t is inflation and D_t is a proxy for monetary policy. Then this stylized model captures a direct effect of monetary policy on inflation and an indirect effect on output through the effect of inflation on output. In this case, Granger tests will fail to detect a causal link between monetary policy and output while Sims tests will 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., ε_{Dt}). In contrast, Granger non-causality is formulated as a restriction on the relation between output and all lagged variables, including covariates that themselves have responded to the policy shock of interest. Granger causality therefore provides an incorrect answer to a question that Sims causality tests answer correctly: will output change in response to a random manipulation if we randomly shock monetary policy?

This example raises the question of how important time-varying, policy-sensitive covariates are in practice. In research on monetary policy, Shapiro (1994) and Leeper (1997) argue that it is important to include inflation in the conditioning set when attempting to isolate the causal effect of monetary policy innovations.

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. Tests of (1) on the other hand are difficult to construct unless D_t, Y_t and X_t can be nested within a linear dynamic model such as an SVAR model. 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.

[7] Many authors have studied the relationship between Granger and Sims-type conditional independence restrictions. See, for example, Dufour and 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 are unrealistic for applications in macroeconomics.

In the remainder of the paper, we assume the policy variable of interest is multinomial. This is in the spirit of a line 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)[8] To develop this setup, we assume that models for the policy function can be written in the parametric form $\Pr(D_t|z_t) = p(z_t, \theta_0)$ for some function $p(.,.)$ and an unknown parameter vector, θ_0. There is a direct analogy to SVAR analysis when χ_t has a representation such as in Example 1. In that case, $p(z_t, \theta_0)$ 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.

Under the non-causality null hypothesis it follows that $\Pr(D_t|z_t, Y_{t+1}, ..., Y_{t+j}, ...) = \Pr(D_t|z_t)$. 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.[9] 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

[8] The recent 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 mutli-valued or continuous causal variables like the Federal funds rate. For a recent extension of cross-sectional propensity-score methods to multi-valued treatments, see Hirano and Imbens (2004).

[9] 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 Eliasz (2005).

develops a parsimonious Probit model of Fed decision-making as a function of net present value measures of inflation and unemployment[10]. Importantly, while it is impossible to know for sure whether a given set of conditioning variables is adequate, our framework naturally generates a diagnostic test that can be used to decide when the model for the policy propensity score is consistent with the data. We illustrate the interaction between specification testing and causality testing in Section 5, below.

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, $y_t = (Y'_{t+1}, ..., Y'_{t+m})'$ but it is also possible to focus on particular future outcomes, say, $y_t = Y'_{t+m}$, when causal effects are thought to be delayed by m periods. Assuming that D_t is a discrete variable taking on $\mathcal{M}+1$ distinct values, the conditional independence hypothesis can be written

$$\Pr(y_t \leq y, D_t = i | z_t) = \Pr(y_t \leq y | z_t) \Pr(D_t = i | z_t) \text{ for } i = \{0, 1, ..., \mathcal{M}\}. \tag{4}$$

We use the short hand 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 θ. A convenient representation of the hypotheses we are interested in testing can be obtained by noting that under the null,

$$\Pr(y_t \leq y, D_t = i|z_t) - \Pr(y_t \leq y|z_t) p(z_t) = E\left[\mathbf{1}(y_t \leq y)\left(\mathbf{1}(D_t = i) - p_i(z_t)\right)|z_t\right] = 0 \text{ for } i = \{0, 1, ..., \mathcal{M}\}. \tag{5}$$

It is convenient to write the moment conditions (5) in vector notation. Noting that $\sum_{i=0}^{\mathcal{M}+1} \mathbf{1}(D_t = i) = 1$ and $\sum_{i=0}^{\mathcal{M}+1} p_i(z_t, \theta) = 1$ we define $\mathcal{M} \times 1$ vectors $\mathcal{D}_t = (\mathbf{1}(D_t = 1), ..., \mathbf{1}(D_t = \mathcal{M}))'$ and $p(z_t) = (p_1(z_t), ..., p_{\mathcal{M}}(z_t))'$ such that the $\mathcal{M}$ non-redundant moment conditions of (5) can be expressed as

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

This leads to a simple interpretation of test statistics based on this moment condition as looking for a relation between (generalized) policy innovations, $\mathcal{D}_t - p(z_t)$, and the distribution of future outcomes. Note also that, like the Hirano, Imbens and Ridder (2000) and Abadie (2005) propensity-score-weighted estimators and the Robins, Mark, and Newey's (1992) partially linear estimator, test statistics constructed

[10] 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.

from moment condition (5) work directly with the propensity score; in particular, no matching step or nonparametric smoothing is required once estimates of the score have been constructed.[11]

We now define $U_t = (y_t, z_t)$ so that the null hypothesis of conditional independence can be represented very generally in terms of moment conditions for functions of U_t. Let $\phi(.,.) : \mathbb{R}^{k_1} \times \mathbb{R}^{k_2} \to \mathbb{H}$ 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 $\phi(.,.)$ 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 ϕ are $\phi(U_t, v) = \mathbf{1}\{U_t \leq v\}$ or $\phi(U_t, v) = \exp(iv'U_t)$ where $i = \sqrt{-1}$, as suggested by Bierens (1982) and Su and White (2003). Since correct specification of the policy model implies that $E[(\mathcal{D}_t - p(z_t))|z_t] = 0$ testing (6) is equivalent to testing the unconditional moment condition $E[\phi(U_t, v)(\mathcal{D}_t - p(z_t))] = 0$ over a sufficiently flexible class of functions $\phi(U_t, v)$ such as $\mathbf{1}\{U_t \leq v\}$.

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. Possibilities include $\phi(U_t, v) = y_t \mathbf{1}\{z_t \leq v_2\}$ which could be used to test if the policy innovation affects the mean of y_t. Generalizations to the effects on higher moments can be handled in the same way. For example, if y_t is univariate, the function $\phi(U_t, v) = y_t^r \mathbf{1}\{z_t \leq v_2\}$ can be used to test if the policy innovation affects the r-th moment of the distribution of the outcome variables. A series of tests thus can be designed to distinguish the effects of policy innovations on the mean and variance of the outcome variable.

The fact that D_t takes on distinct values is well suited to analyze the effects of specific policy actions on the outcome variable. In our empirical application, D_t is modelled to represent situations where the Fed raises, lowers or leaves the interest rates unchanged. By focusing on specific cases, such as $E[\phi(U_t, v)(\mathbf{1}(D_t = i) - p_i(z_t))] = 0$ for $i = 1, ...\mathcal{M}$ separately and therefore allowing for the possibility that the non-causality moment condition may be violated only for certain values of i, we can test if raising or lowering the interest rate has a different effect on the outcome variable. Our approach thus allows for general non-linear responses to policy innovations without the need to explicitly model the functional form of these responses.

An implication of (6) is that the average policy effect is zero as well. In other words, letting E_z be the expectation operator integrating over z_t one obtains, from the law of iterated expectations, that

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

[11] Hirano, Imbens and Ridder (2003) show in a somewhat different context that non-parametric estimation of the propensity score may lead to more efficient inference. Based on their insight it is possible that a test based on a non-parametric estimate of the propensity score would be more powerful than our semiparametric test. We do not consider this type of procedure because the sample size in our application does not lend itself to non-parametric estimation of the propensity score.

In practice, the unconditional moment restriction (7) is often of more direct interest than testing for full conditional independence in (6). Partitioning $v = (v_1, v_2)$ where $v_1, v_2 \in \mathbb{R}^k$ one then can restrict attention to tests based on $\phi(y_t, v_1) : \mathbb{R}^{k_1} \to \mathbb{H}$. In addition it is often the case in applications, as our empirical work in Section 5 illustrates, that the case where y_t is scalar is of most interest.

A third version of our test concerns specification tests for the policy model. Under correct specification of $p(z_t)$ the conditional moment restriction $E\left[(\mathcal{D}_t - p(z_t))|z_t\right] = 0$ must hold. Choosing functions $\phi(z_t, v_2) : \mathbb{R}^{k_2} \to \mathbb{H}$ which are sufficiently flexible, the conditional moment restriction is equivalent to the unconditional moment restriction

$$E\left[\phi(z_t, v_2)(\mathcal{D}_t - p(z_t))\right] = 0. \tag{8}$$

In our empirical application we use tests based on (8) to validate our empirical specification of $p\left(z_t\right)$.

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. In particular, the moment condition in (5) implies that for any fixed y, $\mathbf{1}\left(y_t \le y\right)\left(\mathcal{D}_t - p(z_t)\right)$ is a MDS. Our test is a joint test of whether the set of all processes indexed by $y \in \mathbb{R}^{k_1}$ have the MDS property. We use the terminology of a functional martingale difference hypothesis to distinguish the hypothesis being tested here from the simple MDS hypothesis usually covered in the literature. The functional MDS hypothesis is an extension of the case analyzed by Koul and Stute (1999). The functional nature of the MDS hypothesis related to tests of (6) implies that the test statistic depends on the parameter $v \in \mathbb{R}^k$ where for k it is necessary that $k \ge 2$ while Koul and Stute only consider the case $k = 1$.[12]

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

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

with

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

Under regularity conditions that include stationarity of the observed process, we show in Appendix A that $V_n(v)$ converges weakly to a limiting mean-zero Gaussian process $V(v)$ on the space of cadlag functions[13] denoted by $\mathfrak{D}\left[-\infty, \infty\right]^k$ with covariance function $\Gamma(v, \tau)$, defined as

$$\Gamma(v, \tau) = E\left[V_n(v)V_n(\tau)'\right]$$

[12] Another important difference is that in our setup, the process $\mathbf{1}\left(y_t \le y\right)\left(D_t - p(z_t)\right)$ is not Markovian even under the null hypothesis. This implies that the proofs of Koul and Stute do not apply directly for our case.

[13] Cadlag functions are functions which are continuous from the right with left limits.

where $\nu, \tau \in \mathbb{R}^k$.[14] Using the fact that under the null $E[\mathcal{D}_t|z_t, y_t] = E[\mathcal{D}_t|z_t] = p(z_t)$ and partitioning $u = (u_1, u_2)$ with $u_2 \in [-\infty, \infty]^{k_2}$ we define $H(v \wedge \tau)$ with

$$H(v) = \int_{-\infty}^{v} \left(\operatorname{diag}\left(p(u_2)\right) - p(u_2)p\left(u_2\right)'\right) dF_u\left(u\right) \tag{9}$$

where $\operatorname{diag}\left(p(u_2)\right)$ is the diagonal matrix with diagonal elements $p_i\left(z_t\right)$, $F_u(u)$ is the cumulative marginal distribution function of U_t and $\wedge$ denotes the element by element minimum. The covariance function $\Gamma(v, \tau)$ can now be written as $\Gamma(v, \tau) = \int \phi(u, v) dH\left(u\right) \phi(u, \tau)'$. Note that if $\phi(U_t, v) = \mathbf{1}\left\{U_t \leq v\right\}$ then $\Gamma(v, \tau) = H(v \wedge \tau)$. This is the case we consider in the empirical application. Let $\|m\|^2 = \operatorname{tr}\left(mm'\right)$ 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 KS$= \sup_v \|V_n\left(v\right)\|$ or

$$\mathrm{VM} = \int \|V_n\left(v\right)\|^2 dF_u(v) \tag{10}$$

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 $V_n(v)$ on $\phi\left(U_t, v\right)$, 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 $\Gamma\left(v, \tau\right)$. 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 ϕ, 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 θ needs to be estimated. We use the notation $\hat{V}_n(v)$ to denote test statistics that are based on an estimate $\hat{\theta}$ for θ. 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 θ. The resulting 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 $\hat{V}_n(v)$.

[14] 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 martinagle transformations in Section 4.1 rely on Gaussian limit distributions. Park and Phillips 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 mulitnomial choice case and apply it to the fed funds target rate. The quesiton of how to adapt these results to the problem of conditional independence testing is left for future work.

The combined application of the Rosenblatt and Khmaladze transforms that we advocate in this paper 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:

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

A more formal statement of this assumption is contained in Condition 8 in Appendix A. In our context, $l\left(D_t, z_t, \theta\right)$ 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 $\hat{V}_n\left(v\right)$ induced by the estimation of θ, define the function $\bar{m}(v, \theta) = E\left[m(y_t, D_t, z_t, \theta; v)\right]$ and let

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

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

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

with $L\left(\theta_0\right) = E\left[l\left(D_t, z_t, \theta_0\right) l\left(D_t, z_t, \theta_0\right)'\right]$ as shown in Appendix A. Next we turn to details of the transformations. Section 4.1 discusses a Khmaladze-type martingale transformation that corrects $\hat{V}\left(v\right)$ for the effect of estimation of θ. 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 $T\hat{V}\left(v\right)$ with the property that the transformed process, $W(v) = T\hat{V}\left(v\right)$, is a mean zero Gaussian process with covariance function $\Gamma(v, \tau)$. While $\hat{V}\left(v\right)$ has a complicated data-dependent limiting distribution (because of the estimated θ), 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 $n^{-1/2}\sum_{t=1}^{n} l\left(D_t, z_t, \theta_0\right)$ this process does not have independent increments. Defining $\mathcal{F}_v = \sigma\left(\tilde{V}(s), s \leq v\right)$, we can understand the Khmaladze transform as being based on the insight that, because $\hat{V}(v)$ is a Gaussian process, $\Delta W(v) = \Delta\hat{V}(v) - E\left(\Delta\hat{V}(v)\,|\mathcal{F}_v\right)$ 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 $A_{-\infty} = \varnothing$, $A_\infty = [-\infty,\infty]^k$, $\lambda \leq \lambda' \Longrightarrow A_\lambda \subset A_{\lambda'}$ and $A_{\lambda'} \backslash A_\lambda \to \varnothing$ as $\lambda' \downarrow \lambda$. Define the projection $\pi_\lambda f(v) = \mathbf{1}\left(v \in A_\lambda\right) f(v)$ and $\pi_\lambda^\perp = 1 - \pi_\lambda$ such that $\pi_\lambda^\perp f(v) = \mathbf{1}\left(v \notin A_\lambda\right) f(v)$. We then define the inner product $\langle f(.), g(.)\rangle := \int f(u)' dH(u)\, g(u)$ and, for

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

define the matrix

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

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 $V(\phi(.,v)) = V(v) = \int \phi(u,v) db(u)$. Using the same notation the transformed statistic $W(v)$ is given by

$$T\hat{V}(v) := W(v) = \hat{V}(v) - \int \left\langle \phi(.,v)', d\left(\pi_\lambda \bar{l}(.,\theta)\right)\right\rangle C_\lambda^{-1} \hat{V}(\pi_\lambda^\perp \bar{l}(.,\theta)') \tag{11}$$

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

We show in Appendix A 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) 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. [15]

[15] Stute, Thies and Zhu (1998) analyze a test of conditional mean specification in an independent sample allowing for

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 $l(.,\theta_0)$. 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_λ 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 where

$$\hat{W}_n(v) = T_n V_n(v) = \hat{V}_n(v) - \int \left(\int \phi(u,v) d\hat{H}_n(u) d\left(\pi_\lambda \bar{l}(u,\theta)\right)\right) \hat{C}_\lambda^{-1} \hat{V}_n(\pi_\lambda^\perp \bar{l}(.,\hat{\theta})') \tag{12}$$

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

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

$$A_\lambda = [-\infty, \lambda] \times [-\infty, \infty]^{k-1}, \tag{13}$$

which lead to test statistics with simple closed form expressions. Denote the first element of y_t by y_{1t}. Then (12) can be expressed more explicitly as

$$\hat{W}_n(v) = \hat{V}_n(v) - n^{-1/2} \sum_{t=1}^n \left[\phi\{U_t, v\} \frac{\partial p(z_t,\hat{\theta})}{\partial \theta'} \hat{C}_{y_{1t}}^{-1} n^{-1} \sum_{s=1}^n \mathbf{1}\{y_{1s} > y_{1t}\} \bar{l}(U_s,\hat{\theta})' \left(\mathcal{D}_s - p(z_s,\hat{\theta})\right)\right] \tag{14}$$

In the appendix 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 $U_t = [U_{t1}, ..., U_{tk}]$ and define the transformation $w = T_R(v)$ component wise by $w_1 = F_1(v_1) = \Pr(U_{t1} \leq v_1)$, $w_2 = F_2(v_2|v_1) = \Pr(U_{t2} \leq v_2 | U_{1t} = v_1), ..., w_k =$

heteroskedasticity by rescaling the equivalent of our $m(y_t, D_t, z_t, \theta_0; v)$ by the conditional variance. But their approach does not work for our problem because the relevant conditional variance depends on the unknown parameter θ. Instead of correcting $m(y_t, D_t, z_t, \theta_0; v)$ we adjust the transformation T in the appropriate way.

$F_k(v_k|v_{k-1},...,v_1)$ where $F_k(v_k|v_{k-1},...,v_1) = \Pr(U_{tk} \leq v_k|U_{tk-1} = v_{k-1},...,U_{t1} = v_1)$. The inverse $v = T_R^{-1}(w)$ of this transformation is obtained recursively as $v_1 = F_1^{-1}(u_1)$,

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

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

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

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

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

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

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

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

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. Of course, our test statistic also depends on the choice of $\phi(.,.)$. 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[16].

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 θ with $\hat{\theta}$ in $V_{w,n}(v)$. Define the transform $T_w\hat{V}_w(w)$ as before by[17]

$$T_w\hat{V}_w(w) := W_w(w) = \hat{V}_w(w) - \int\left\langle\phi(.,w)', d\pi_\lambda\bar{l}_w(.,\theta)\right\rangle C_\lambda^{-1}\hat{V}_w(\pi_\lambda^\perp\bar{l}_w(.,\theta)'). \tag{15}$$

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 $W_w(\phi(.,w)) = W_w(w)$ to emphasize the dependence of W on ϕ, it follows from the previous discussion that

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

[16] In the working paper (Angrist and Kuersteiner, 2004) we discuss ways to resolve the problem of the ordering in w_t.

[17] For a more detailed derivation see Appendix B.

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

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

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) of Section A.2. A simple way of constructing higher order kernels is given in Bierens (1987). Let $K_k(x) = (2\pi)^{-k/2}\sum_{j=1}^{\omega}\theta_j\left|\sigma_j\right|^{-k}\exp\left(-1/2x'x/\sigma_j^2\right)$ with $\sum_{j=1}^{\omega}\theta_j = 1$ and $\sum_{j=1}^{\omega}\theta_j\left|\sigma_j\right|^{2\ell} = 0$ for $\ell = 1,2,...,\omega-1$. Let $m_n = O(n^{-1/(2+k)})$ be a bandwidth sequence and define

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

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

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

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

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

The empirical version of the transformed statistic is

$$\begin{aligned}
\hat{B}_{\hat{w},n}(w) &= \hat{W}_{\hat{w},n}\left(\phi(.,w)\hat{h}_w(.)^{-1/2}\right) \\
&= n^{-1/2}\sum_{t=1}^{n}\phi(\hat{w}_t,w)\,\hat{h}(z_t)^{-1/2}\left[D_t - p(z_t,\hat{\theta}) - \hat{A}_{n,t}\right] \qquad (16)
\end{aligned}$$

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 7 in Appendix A 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. Theorem 7 shows that $\hat{B}_{\hat{w},n}(w) \Rightarrow B_w(w)$ on $\mathfrak{D}\left[\Upsilon_{[0,1]}\right]$ where $B_w(w)$ is a standard Gaussian process and $\Upsilon_{[0,1]} = \left\{w \in [0,1]^k \,|w = \pi_x w\right\}$ where $\pi_x w = \mathbf{1}(w \in A_x) w$ for $x \in \mathbb{R}$ and A_x is the set defined in (13). 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

$$\mathrm{VM}_w = \int_{\Upsilon_{[0,1]}} \left\|\hat{B}_{\hat{w},n}(w)\right\|^2 dw \Rightarrow \int_{\Upsilon_{[0,1]}} \|B_w(w)\|^2 dw \tag{17}$$

and

$$\mathrm{KS}_w = \sup_{v \in \Upsilon_{[0,1]}} \left\|\hat{B}_{\hat{w},n}(w)\right\| \Rightarrow \sup_{v \in \Upsilon_{[0,1]}} \|B_w(w)\|. \tag{18}$$

Here VM_w and 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 (17) and the supremum in (18) can be computed over a discrete grid. The asymptotic representations (17) and (18) make it possible to use asymptotic statistical tables. For the purposes of the empirical application below, we computed critical values for the VM statistic in the special case where $\phi(.,v) = \mathbf{1}\{. \leq v\}$ 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 $s = 1, ..., S$ bootstrap replications according to

$$\varepsilon_{t,s}^* = \varepsilon_{t,s}^{**}/\sqrt{2} + \left(\left(\varepsilon_{t,s}^{**}\right)^2 - 1\right)/2 \tag{19}$$

where $\varepsilon_{t,s}^{**} \sim N(0,1)$ is independent of the sample. Let the moment condition underlying the transformed test statistic (16) be denoted by

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

and write

$$\hat{B}^{*}_{\hat{w},n;s}(w) = n^{-1/2}\sum_{t=1}^{n}\varepsilon^{*}_{t,s}\left(m_{T,t}\left(v,\hat{\theta}\right) - \bar{m}_{n;T}\left(v,\hat{\theta}\right)\right) \tag{20}$$

to denote the test statistic in a bootstrap replication, with $\bar{m}_{n;T}\left(v,\hat{\theta}\right) = n^{-1}\sum_{t=1}^{n} m_{T,t}\left(v,\hat{\theta}\right)$. The distribution of $\varepsilon^{*}_{t,s}$ induced by (19) guarantees that the first three empirical moments of $m_{T,t}\left(v,\hat{\theta}\right) - \bar{m}_{n;T}\left(v,\hat{\theta}\right)$ are preserved in bootstrap samples. Theorem 8 in the appendix 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 $s = 1, ...S$ samples $\varepsilon^{*}_{1,s}, ..., \varepsilon^{*}_{n,s}$ independently from the distribution (19); 2) compute $\mathrm{VM}_s = \int_{\Upsilon_{[0,1]}}\left\|\hat{B}^{*}_{\hat{w},n;s}(w)\right\|^2 dw$ for $s = 1, ..., S$; 3) obtain the desired empirical quantile from the distribution of VM_s, $s = 1, ..., S$. 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

$$\hat{V}^{*}_{n;s}(v) = n^{-1/2}\sum_{t=1}^{n}\varepsilon^{*}_{t,s}\left(m(y_t, D_t, z_t, \hat{\theta}; v) - \bar{m}_n(v,\hat{\theta})\right) \tag{21}$$

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 using 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 are 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 Romer's (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.[18] In much of the period since the mid-1970s, and especially in the Greenspan era,

[18] 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.

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 the decisions about whether and how to redirect policy is available to researchers studying the effects of monetary policy. 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 Eliasz (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. Identification assumptions about the transmission mechanism of policy innovations are then required to disentangle the effects of monetary policy.

Our approach is closer to the recursive identification strategy employed by Christiano, Eichenbaum, and Evans (1999), hereafter CEE. The CEE study similarly 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 both endogenous because of an immediate effect of monetary policy shocks on producer prices. In contrast, the *intended* funds rate is determined by forecasts of market conditions and intentions formed by policy makers only based on predetermined information, and thus is sequentially exogenous by construction. Moreover, the CEE approach is parametric and relies on linear models for both outcomes and policy.

The substantive identifying assumption in our framework (and 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 what the best model for the policy propensity score is (.e.g., there is some uncertainly as to the flexibility and lag length). We explore these issues by experimenting 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. The monthly sample includes 29% reductions, 32% increases and 39% no change in the intended funds rate.[19] Following Hamilton and Jorda (2002), we fit ordered probit models with dDff_t as the dependent variable. Hamilton and Jorda (2002) derive their specification using a linear latent-index model of the central bank's intentions.

The first specification we report on, labeled 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.[20] 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 increments. 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 model (b). To allow for non-linear dynamic responses, model (c) adds a quadratic function of past intended changes in the federal funds rate to the restricted model (b). We also consider versions of (a)-(c) using a discretized variable for the lagged change in the intended federal funds rate. These 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.[21] This baseline Taylor specification is labeled model (g). We also consider a modification replacing dff_{t-2} with $(\mathrm{dff}_{t-1})^2$ to capture non-linearities (model h). Finally, we look at variants (i) and (j) of (g) and (h), that replace lags of dff_t with the corresponding lags of dDff_t.

As a benchmark for our semiparametric analysis, our analysis begins 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

[19] 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 were treated as separate observations.

[20] The unemployment innovation is the Romer's $\tilde{u}_{m0}$, the Greenbook forecast for the unemployment rate in the current quarter, minus the unemployment rate in the previous month.

[21] Monthly GDP is interpolated from quarterly using a program developed by Mönch and Uhlig (2005). We thank Emanuel Mönch and Harld 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.

with future outcome variables. Under the null hypothesis of no causal effect, future outcome variables should have insignificant coefficients in the policy model.[22] This is the essence of Condition 2.

Tables 1a and 1b report results for parametric Sims tests for the effect of policy on the non-seasonally-adjusted index of industrial production. The table shows tests for the effect on the cumulated change in industrial production up to three years ahead. This is presented at quarterly, half year and full year intervals. 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 and more quarters lead. This result is robust across models (d)-(f) and (i)-(j). There is also some isolated significance at the 10% level at earlier leads for models (e)-(j). For models with dff_t on the right hand side, the lag pattern is more mixed. Models (a),(b) and (h) predict a response after 7 quarters, while model (c) predicts a response after 8 quarters and model (g) predicts a response after 6 quarters. We note that model (h) generates an isolated initial impact of the monetary policy shock, but this does not persist in the coarser half and full year tests. Tests at the 10 percent level generally show earlier effects, 6-7 quarters out for 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 that does not require specification of an elaborate model for the response function. Importantly, the semiparametric tests can also be used to evaluate the policy model specification and explore possible misspecification due to functional form or omitted covariates. This is done by using the moment condition in (8) with the function $\phi(z_t, v_2)$ configured to equal $\mathbf{1}\{z_{ti} \leq v_{2i}\}$ and i ranging across all covariates of models (a) through (j). Rejection of the test statistic based on (8) implies that the policy model is misspecified.

The results of specification tests for the policy models appear in Tables 2a and 2b. Specifically, the table shows test statistics for possibly omitted covariates (or lagged covariates), and starred significance levels (*=10%, **=5% and ***=1%) for each covariate-specific test in each model. The table reports test results using asymptotic critical values (ASY), the bootstrap defined in (20, BSK) and the bootstrap for the untransformed statistic (21, BS).[23] Tests based on (20) should have the most accurate size since the test statistic is asymptotically pivotal.

The best fit in the specification tests, as determined by BSK, is the baseline Romer model (a) as well as

[22] See also the discussion at the end of Section 2.

[23] 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 independetly $\varepsilon_{t,i}$ from an iid standard normal distribution. The sample size was set to $n = 300$ and 100,000 replications were done. We then compute $B_i^{**}(w) = n^{-1/2}\sum_{t=1}^{n} \varepsilon_t^{*}\mathbf{1}\left\{U_{t,i}^{**} \leq x\right\}$ for each replication sample i. The final step consists in forming the sum $B_i^{**} = n^{-1}\sum_{t=1}^{n}\mathbf{1}\left\{U_{j,i}^{**} \in \Upsilon_{[0,1]}\right\}\left\|B_i^{**}\left(U_{j,i}^{**}\right)\right\|^2$. Asymptotic critical values are then obtained from the quantiles of the empirical distribution of B_i^{**}.

modifications (c) and (e). The Taylor models generally fit less well, with moment restrictions violated most notably for the innovation in the Greenbook forecast for the percentage change in GDP/GNP variables. 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 on the other hand seem to implicitly take account of lagged real GDP growth and inflation, in spite of the fact that these variables are not included in the Romer propensity score. Of the Romer models (a), (c) and (e) look best, while for the Taylor models, specification (h) seems best.

We now turn to the semiparametric causality tests based on the unconditional moment conditions in (7) with $\phi(U_t, v)$ configured to equal $\mathbf{1}\{y_t \leq v_1\}$. 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 2b, these tables report test statistics and starred significance levels. Results using ASY and BSK are generally similar, though ASY appears to reject slightly more often. This is particularly true for the 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 (21) tend to reject much less often in most models, indicating some undersizing for these tests, especially in light of the parametric tests in Tables 1a and 1b.

The multivariate tests look simultaneously at the significance of up and down movements in a single test statistic, in a manner analgous to the parametric tests in table 1. All specifications in Table 3 show significant effects at the 5% level starting 10 quarters ahead. Model (a) also generates significant effects as early as in quarter 7, using both asymptotic and bootstrap critical values. The Taylor models (h), (i) and (j) also generate significant effects starting in quarter 8 using both ASY and BSK. The restricted Romer models, (b), (c), (e) and (f), generate the longest lag in policy effects at about 10 quarters, although models (e) and (f) also show weaker significance at the 10% level as early as 4 quarters ahead in model (e) and 3 quarters ahead in model (f).

We also considered the effects of positive and negative monetary shocks separately. The asymmetric tests again use the moment conditions (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 expectational surprise variable at the core of our testing framework. To save space, the asymmetric results are shown only for models that do well in the model specification tests in Table 2. These are (a), (c), and (e) for the Romer specifications and (h) for the Taylor model.

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% ciritcal level are (a) and (c), where a response appears after 10 quarters. (The ASY and BSK results are similar on this point). Results for Taylor model, (h), generate an isolated significant test two-and-a-half years out, though only when using the bootstrap critical values, BSK. There is a less significant (10% level) response in models (e) and (h) at a 10-11 quarter lead as well.

The results in Table 5 contrast sharply 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. Model (c) also shows a significant effect, but only for critical values based on BS. For models (a) and (h) we also find a less significant early response at 4 and 5 quarters. The effects are generally more significant in Table 5 than in Table 4. This is evident both from the fact that the BS-based test generates a significant result and from the fact that some results are significant at the 1% level using both the ASY and BSK tests. Overall, test results using ASY and BSK are similar, except for the quarterly tests of the Taylor model where the BSK based tests indicate slightly more significance at shorter leads.

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 somewhat more immediate response for the Romer based specifications (a) than for the Taylor based 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 enogenous 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 alternatives of interest, 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 shock after about 7 quarters, while the Taylor model and a restricted Romer specification shows responses that take a little longer to develop. Our results are broadly in line with those in Romer and Romer (2004), who find the strongest response to a monetary shock after about 2 years with continued significance for another year. Our results therefore highlight the robustness of the Romers' original findings. An investigation that allows for different reponses to rate increases and decreases shows an earliy and strong response to increases without much response to decreases. This finding has not featured in most previous discussions of the causal effects of monetary shocks.

Finally, in contrast with the Romer's findings and those reported here, the SVAR literature generally finds somewhat less significant and 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 rapdily declining hump shaped response. Thus, while SVAR findings similarly suggest that monetary policy matters, some of the early impact in the SVAR literature may be generated in part by the structural assumptions used to identify these models.

A Asymptotic Critical Values

This Appendix provides formal results on the distribution of the test statistics described above and forms the basis for the construction of asymptotic critical values. The theorems and proofs use the additional notation outlined below.

A.1 Additional Notation and Assumptions

We focus initially on the process $V_n(v)$ and the associated transformation T. Results for $V_{w,n}(w)$ and the transformed process $T_w V_{w,n}(w)$ then follow as a special case.

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 $\left(\mathbb{R}^{k+1}, \mathcal{B}^{k+1}\right)$ where $\mathcal{B}^{k+1}$ is the Borel σ-field on $\mathbb{R}^{k+1}$ and k is fixed with $2 \leq k < \infty$. Let $\mathcal{A}_1^l = \sigma(\chi_1, ..., \chi_l)$ be the sigma field generated by $\chi_1, ..., \chi_l$. The sequence χ_t is β-mixing or absolutely regular if

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

A sequence is called α-mixing if

$$\alpha_m = \sup_{l \geq 1} E\left[\sup_{A \in \mathcal{A}_1^l, B \in \mathcal{A}_{l+m}^{\infty}} \left|\Pr(A \cap B) - \Pr(A)\Pr(B)\right|\right] \to 0 \text{ as } m \to \infty$$

and it is well known that $\alpha_m \leq \beta_m$.

Condition 3 *Let χ_t be a stationary, absolutely regular process such that for some $2 < p < \infty$ the β-mixing coefficient of χ_t satisfies $m^{p/(p-2)} (\log m)^{2(p-1)/(p-2)} \beta_m \to 0$.*

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

Condition 5 *The matrix of functions $\phi(.,.)$ belongs to a VC subgraph class of functions with envelope $M(\chi_t)$ such that $E\|M(\chi_t)\|^{2+\delta} < \infty$ for some $\delta > 0$.*

We note that $|m(y_t, D_t, z_t, \theta_0|v)| \leq 2$ for $\phi(., v) = \mathbf{1}\{. \leq v\}$ such that by Pollard (1984) Theorem II.25, $m_v(W_t) = m(y_t, D_t, z_t, \theta_0|v)$ is a VC subgraph class of functions indexed by v with envelope 2.

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, τ 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 $h(v) = \partial^k H(v)/\partial v_1 ... \partial v_k$ and assume that* $\det(h(v)) > 0$ *for all $v \in \mathbb{R}^k$.*

Remark 1 *A sufficient condition for Condition 6 is that* $0 < p_i(z_t, \theta_0) < 1$ *almost surely for all* $i = 0, 1, ..., \mathcal{M}$.

A.2 Limiting Distributions

Let $\mathfrak{D}[-\infty,\infty]^k$ be the space of functions that are continuous from the right with left limits (Cadlag) mapping $[-\infty,\infty]^k \to \mathbb{R}$. We consider weak convergence on $\mathfrak{D}[-\infty,\infty]^k$ equipped with the sup norm. Here $[-\infty,\infty]^k$ denotes the k-fold product space of the extended real line equipped with the metric $q(v,\tau) = \left(\sum_{i=1}^k |\Phi(v_i) - \Phi(\tau_i)|^2\right)^{1/2}$ where Φ is a fixed, bounded and strictly increasing function. It follows that $[-\infty,\infty]^k$ is totally bounded. The function space $\mathcal{F} = \left\{m(.,v)|v \in [-\infty,\infty]^k\right\}$ of functions m indexed by v then is a subset of the space of all bounded functions on $[-\infty,\infty]^k$ denoted by $l^\infty([-\infty,\infty]^k)$.

Proposition 2 *Assume that Conditions 3, 4 and 6 are satisfied. Let* $v_i \in [-\infty,\infty]^k$ *for* $i = 1, ..., s$ *be a finite collection of points. Then, for all finite* s, $V_n(v_1), ..., V_n(v_s)$ *converges in distribution to a Gaussian limit with mean zero and covariance function* $\Gamma(v_i, v_j)$. *Moreover,* $V_n(v)$ *converges in* $\mathfrak{D}[-\infty,\infty]^k$ *to a Gaussian process* $V(v)$ *with covariance kernel* $\Gamma(v,\tau)$ *with* $v, \tau \in [-\infty,\infty]^k$ *and* $V(-\infty) = 0$, $H(v)$ *is positive definite with* $H(v)$ *increasing in* v.

Proof of Proposition 2. As noted before, under H_0, $m_v(\chi_t)$ is a martingale difference sequence such that $E(m_v(\chi_t)|z_t) = 0$. Let $\lambda = (\lambda_1, ..., \lambda_s)'$ with $\|\lambda\| = 1$ and $\lambda_i \in \mathbb{R}^{\mathcal{M}}$. For finite dimensional convergence we apply Corollary 3.1 of Hall and Heyde (1980) to $Y_t = \lambda_1' m_{v_1}(\chi_t) + \lambda_2' m_{v_2}(\chi_t) + ... + \lambda_s' m_{v_s}(\chi_t)$. Then, clearly Y_t is also a martingale difference sequence. Consider $Y_{nt} = Y_t/\sqrt{n}$. Then, for all $\varepsilon > 0$,

$$\sum_t E\left(Y_{nt}^2 \mathbf{1}\{|Y_{nt}| \geq \varepsilon\} |\mathcal{A}_1^{t-1}\right) \leq \sum_t E\left(Y_{nt}^2 \mathbf{1}\left\{\|M(\chi_t)\| \textstyle\sum_i \|\lambda_i\| \geq \sqrt{n}\varepsilon\right\} |\mathcal{A}_1^{t-1}\right) \to 0 \text{ a.s.}$$

because $E\|M(\chi_t)\|^{2+\delta}$ is bounded for some $\delta > 0$. Also,

$$\begin{aligned} \sum_t E\left[Y_{nt}^2|\mathcal{A}_1^{t-1}\right] &= n^{-1}\sum_{t=1}^n E\left[Y_t^2|\mathcal{A}_1^{t-1}\right] \\ &= n^{-1}\sum_{t=1}^n \sum_{i,j=1}^s E\left[\lambda_i' \boldsymbol{\phi}(u_t, v_i)\left(\operatorname{diag}(p(z_t)) - p(z_t)p(z_t)'\right)\boldsymbol{\phi}(u_t, v_j)' \lambda_j |\mathcal{A}_1^{t-1}\right] \\ &\xrightarrow{p} \sum_{i,j=1}^s \lambda_i' \Gamma(v_i, v_j)\lambda_j \end{aligned}$$

where the last line is a consequence of Theorem 2.1 in Arcones and Yu (1994). By the Cramer-Wold theorem this establishes finite dimensional convergence. The functional central limit theorem again follows from Theorem 2.1 in Arcones and Yu (1994). ■

The next proposition establishes a linear approximation to the process $\hat{V}_n(v)$ evaluated at the estimated parameter value $\hat{\theta}$. The fact that $l(D_t, z_t, \theta_0)$ is a martingale difference sequence is critical to the development of a distribution free test statistic. The next condition states that the propensity score $p(z_t, \theta)$ is the correct parametric model for the conditional expectation of D_t and lists a number of additional regularity conditions.

Condition 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 $\theta \in \{\theta \in \Theta| \|\theta - \theta_0\| \leq \delta\} := N_\delta(\theta_0)$ 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*

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

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

$$E\left[p_i(z_t, \theta_0)^{-(4+\delta)}\right] < \infty$$

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 $\delta > 0$.

Remark 2 *By Pakes and Pollard (1989, Lemma 2.13) the uniform Lipschitz condition for the derivatives $\partial p_i(z_t|\theta)/\partial\theta_j$ guarantees that the functions $\partial p(z_t|\theta)/\partial\theta'$ indexed by θ form a Euclidean class for the envelope $B(z_t)\left(2\sqrt{d}\sup_{N(\theta_0)}\left\|\theta - \theta'\right\|\right)^\alpha + |\partial p_i(z_t|\theta_0)/\partial\theta_j|$.*

Condition 8 *Let*

$$l(D_t, z_t, \theta) = \Sigma_\theta^{-1} \frac{\partial p'(z_t, \theta)}{\partial\theta} h(z_t, \theta)^{-1} (\mathcal{D}_t - p(z_t, \theta)) \tag{22}$$

where

$$h(z_t, \theta) = \left(\operatorname{diag}(p(z_t, \theta)) - p(z_t, \theta) p(z_t, \theta)'\right)$$

and

$$\Sigma_\theta = E\left[\frac{\partial p'(D_t|z_t, \theta)}{\partial\theta} h(z_t, \theta)^{-1} \frac{\partial p(D_t|z_t, \theta)}{\partial\theta'}\right]. \tag{23}$$

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

Remark 3 *Note that for* $P(z_t,\theta) = \operatorname{diag}(p(z_t,\theta))$ *it follows that*

$$h(z_t,\theta)^{-1} = P(z_t,\theta)^{-1} + \frac{P(z_t,\theta)^{-1} p(z_t,\theta) p(z_t,\theta)' P(z_t,\theta)^{-1}}{\left(1 - p(z_t,\theta)' P(z_t,\theta)^{-1} p(z_t,\theta)\right)} = P(z_t,\theta)^{-1} + \frac{\mathbf{1}\mathbf{1}'}{1-\sum_{j=1}^{\mathcal{M}} p_i(z_t,\theta)}. \quad (24)$$

Simple algebra then shows that

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

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

Proposition 3 *Assume that Conditions 3, 4,5, 6, 7 and 8 are satisfied. Then*

$$\sup_{v\in[-\infty,\infty]^k} \left\| \hat{V}_n(v) - V_n(v) + \dot{m}(v,\theta_0) n^{-1/2} \sum_{t=1}^{n} l(D_t,z_t,\theta_0) \right\| = o_p(1) \quad (25)$$

and if $l(D_t,z_t,\theta_0)$ *is as defined in 22 and 23 then* $\hat{V}_n(v)$ *converges weakly in* $\mathfrak{D}[-\infty,\infty]^k$ *equipped with the sup norm to a limiting Gaussian process with mean zero and covariance function* $\hat{\Gamma}(v,\tau) = \Gamma(v,\tau) - \dot{m}(v,\theta_0)L(\theta_0)\dot{m}(\tau,\theta_0)'$ *where* $L(\theta_0) = \Sigma_{\theta_0}^{-1}$ *is defined in 23.*

Proof of Proposition 3. Note that $\hat{V}_n(v) - V_n(v) = n^{-1/2}\sum_t^n \phi(U_t,v)\left[p(z_t,\theta_0) - p(z_t,\hat{\theta})\right]$ such that we can approximate

$$\begin{aligned}\hat{V}_n(v) - V_n(v) &= \frac{1}{n}\sum_t^n \left(\phi(U_t,v)\left[\frac{\partial p(z_t,\theta_n)}{\partial\theta'} - \frac{\partial p(z_t,\theta_0)}{\partial\theta'}\right]\right)\left(n^{1/2}\left(\hat{\theta}-\theta_0\right)\right) \\ &\quad + \frac{1}{n}\sum_t^n \left(\phi(U_t,v)\frac{\partial p(z_t,\theta_0)}{\partial\theta}\right)\left(n^{1/2}\left(\hat{\theta}-\theta_0\right)\right)\end{aligned}$$

where $\|\theta_n - \theta_0\| \le \left\|\hat{\theta}-\theta_0\right\|$ by the mean value theorem. Let $\dot{m}(\theta,v) = E\left[\phi(U_t,v)\frac{\partial p(z_t,\theta)}{\partial\theta'}\right]$ and $\dot{m}(U_t,\theta,v) = \phi(U_t,v)\frac{\partial p(z_t,\theta)}{\partial\theta'} - \dot{m}(\theta,v)$. From Pakes and Pollard (1989, Lemmas 2.13 and 2.14) and Condition 7 it follows that $\dot{m}(.,\theta,v)$ is a matrix of funcitions in a Euclidean class indexed on $N(\theta_0)\times[-\infty,\infty]^k$ with envelope $\mathcal{M}B(z_t)\left(2\sqrt{d}\sup_{N(\theta_0)}\|\theta-\theta'\|\right)^\alpha + \sum_{i=1}^{\mathcal{M}} |\partial p_i(z_t|\theta_0)/\partial\theta_j|)M(\chi_t)$ for all the elements in the j-th column of $\dot{m}(U_t,\theta,v)$. Note that the factor $\mathcal{M}$ can be replaced with the constant 1 if $\phi(U_t,v)$ is scalar valued. Then

$$\begin{aligned}&\left\|\frac{1}{n}\sum_t^n \phi(U_t,v)\left[\frac{\partial p(z_t,\theta_n)}{\partial\theta} - \frac{\partial p(z_t,\theta_0)}{\partial\theta}\right]\right\| \\ \le\ & \sup_{\|\theta-\theta_0\|\le\delta}\sup_v \left\|\frac{1}{n}\sum_t^n [\dot{m}(U_t,\theta,v) - \dot{m}(U_t,\theta_0,v)]\right\| + \sup_{\|\theta-\theta_0\|\le\delta}\|\dot{m}(\theta,v) - \dot{m}(\theta_0,v)\| + o_p(1) = o_p(1)\end{aligned}$$

since $\sup_{\|\theta-\theta_0\|\leq\delta}\sup_v \left\|\frac{1}{n}\sum_t^n\left[\dot{m}(U_t\theta,v)-\dot{m}(U_t,\theta_0,v)\right]\right\| = o_p(1)$ by applying Lemma 2.1 of Arcones and Yu (1994) to each element $\sup_{\|\theta-\theta_0\|\leq\delta}\sup_v \left|\frac{1}{n}\sum_t^n\left[\dot{m}_{i,j}(U_t\theta,v)-\dot{m}_{i,j}(U_t,\theta_0,v)\right]\right|$. This completes the proof of 25.

The second part of the result follows from the fact that the class of functions $\mathcal{F}=[m_v(.)]_i+[\dot{m}(\theta,v)\,l(.,.,\theta_0)]_i$ where $[.]_i$ denotes the i-the element of a vector, is a Euclidean class by Lemma 2.14 of Pakes and Pollard (1989). Since $m_v(X_t)+\dot{m}(\theta,v)\,l(D_t,z_t,\theta_0)$ is a martingale difference sequence with respect to the filtration $\mathcal{A}_1^{t-1}$ finite dimensional convergence to a Gaussian random vector with zero mean and covariance function $\hat{\Gamma}(v,\tau)$ follows from the martingale CLT (Hall and Heyde, Corollary 3.1) and the fact that $0<\|\Sigma_{\theta_0}\|<\infty$ by Condition 8. Convergence to a weak limit in $\mathfrak{D}\left[-\infty,\infty\right]^k$ then follows again by Theorem 2.1 of Arcones and Yu (1994) as well as van der Vaart and Wellner (1996, Corollary 1.4.5) together with Pakes and Pollard (1989, Lemmas 2.13 and 2.15) to handle the vector case. ■

We now establish that the process $T\hat{V}(v)$, defined in (11) is zero mean Gaussian with covariance function $\Gamma(v,\tau)$. This establishes that the process $T\tilde{V}(v)=W(v)$ can be transformed to a distribution free process via Lemma 3.5 and Theorem 3.9 of Khmaladze (1993).

In order to define the transform T we choose a grid $-\infty=\lambda_0<\lambda_1<...<\lambda_N=\infty$ on $[-\infty,\infty]$, let $\Delta\pi_{\lambda_i}=\pi_{\lambda_{i+1}}-\pi_{\lambda_i}$ and set

$$c_N(V)=\sum_{i=1}^{N}\left\langle\phi\left(.,v\right),\Delta\pi_{\lambda_i}\bar{l}(.,\theta)\right\rangle C_{\lambda_i}^{-1}V(\pi_{\lambda_i}^{\perp}\bar{l}(\vartheta,\theta)). \tag{26}$$

This construction is the same as in Khmaladze (1993) except that we work on $[-\infty,\infty]$ rather than $[0,1]$. In Proposition (4) we show that $c_N(V)$ converges as $N\to\infty$ and $\max_i\left(\Phi(\lambda_{i+1})-\Phi\left(\lambda_i\right)\right)\to 0$. Let the limit of $c_N(V)$ be denoted as $c(V)=\int\left\langle\phi\left(.,v\right),d\pi_\lambda\bar{l}(.,\theta)\right\rangle C_\lambda^{-1}V\left(\pi_\lambda^{\perp}\bar{l}(.,\theta)\right)$

Condition 9 *Let $\{A_\lambda\}$ be a family of measurable subsets of $[-\infty,\infty]^k$, indexed by $\lambda\in[-\infty,\infty]$ such that $A_{-\infty}=\varnothing$, $A_\infty=[-\infty,\infty]^k$, $\lambda\leq\lambda'\implies A_\lambda\subset A_{\lambda'}$ and $A_{\lambda'}\backslash A_\lambda\to\varnothing$ 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 the projection $\pi_\lambda f(v)=\mathbf{1}\left(v\in A_\lambda\right)f(v)$ and $\pi_\lambda^{\perp}=1-\pi_\lambda$ such that $\pi_\lambda^{\perp}f(v)=\mathbf{1}\left(v\notin A_\lambda\right)f(v)$. We then define the inner product $\langle f(.),g(.)\rangle:=\int f(u)dH(u)g(u)'$ and the matrix*

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

Assume that $\langle f(v),\pi_\lambda g(v)\rangle$ is absolutely continuous in λ and C_λ is invertible for $\lambda\in[-\infty,\infty)$.

Proposition 4 *Assume condition 9 holds. Define $\Upsilon_x=\left\{v\in[-\infty,\infty]^k\,|v=\pi_x v\right\}$ for some $x<\infty$. Let $c_N(v)$ be defined as in 26. Then $c_N(v)$ converges with probability 1 to $c(v)$ for all $v\in\Upsilon_x$. Let $T\hat{V}(v)$ be as defined in 11. Then $T\hat{V}(v)$ is a Gaussian process with zero mean and covariance function $\Gamma(v,\tau)$ for all $v,\tau\in\Upsilon_x$.*

Proof of Proposition 4. The proof of this result follows closely Khmaladze (1993) with the necessary adjustments pointed out. First, let $V(v)$ be a Gaussian process on $[-\infty,\infty]^k$ and taking values in $\mathbb{R}^{\mathcal{M}}$ with zero mean and covariance function $\Gamma(v,\tau)$ and $V(-\infty)=0$. See Kallenberg (1997, p. 201) for the construction of such a process. Then, $V(\pi_\lambda^\perp \bar{l}(.,\theta))$ is a process with trajectories that are continuous in λ by essentially the same argument as in Lemma 3.2 of Khmaladze. To see this fix $\alpha \in \mathbb{R}^{\mathcal{M}}$ such that $\alpha' V(\pi_\lambda^\perp \bar{l}(.,\theta))$ is a Wiener process on $[-\infty,\infty]$ with mean zero, $\alpha' V(\pi_\infty^\perp \bar{l}(.,\theta)) = 0$ and variance $\alpha' C_\lambda \alpha$ with almost all trajectories continuous in λ on $[-\infty,\infty]$. To show that $c_N(v) \to c(v)$ almost surely we adapt the proof of Lemma 3.3 of Khmaladze (1993). As there, define $\rho_1(\xi) = |\xi_1| + \dots + |\xi_k|$ for any vector $\xi = (\xi_1, \dots, \xi_k) \in \mathbb{R}^k$ and $\rho_\infty(\xi) = \max_i |\xi_i|$. Set $\xi = \langle \phi, \Delta\pi_\mu \bar{l}(.,\theta) \rangle$ and $\eta(\mu,\lambda) = C_\mu^{-1} V(\pi_\mu^\perp \bar{l}(.,\theta)) - C_\lambda^{-1} V(\pi_\lambda^\perp \bar{l}(.,\theta))$. By Condition 9 the matrix C_λ is invertible on $[-\infty,\infty)$ and C_λ^{-1} is continuous in λ. Then, since $V(\pi_\lambda^\perp \bar{l}(.,\theta))$ is continuous in λ almost surely, we have

$$\sup_{\substack{|\Phi(\lambda)-\Phi(\mu)|<\delta \\ \lambda,\mu\in[-\infty,x]}} \rho_\infty(\eta(\mu,\lambda)) \to 0$$

with probability 1 for any fixed $x<\infty$. The remainder of the proof in Khmaladze (1993) then goes through without change.

We first represent $\hat{V}(v)$ in terms of $V(v)$. Let $V(l(.,\theta_0)) = \int l(u,\theta_0)\,db(u)$ as before for any function $l(v,\theta)$ and $b(v)$ a zero mean vector Gaussian process with covariance function $H(v\wedge\tau)$ and note that $\hat{V}(v) = V(\phi(.,v)) - \dot{m}(v,\theta)\Sigma_\theta^{-1} V(\bar{l}(.,\theta_0)')$. In order to establish a corresponding result to Lemma 3.4 of Khmaladze (1993) we first show that $\hat{V}(v) = V(\phi(.,v)) - \dot{m}(v,\theta)\Sigma_\theta^{-1} V(\bar{l}(.,\theta_0)')$ is a valid representation of the limiting distribution of $\hat{V}_n(v)$ which was derived in Proposition 3. Clearly, $\hat{V}(v)$ is zero mean Gaussian and the covariance function is

$$E\left[V(v)V(\tau)'\right] - \dot{m}(v,\theta_0)\Sigma_\theta^{-1}\int \phi(u,\tau)\,H(du)\bar{l}(u,\theta_0) - \left(\int \phi(u,v)\,H(du)\bar{l}(u,\theta_0)\right)\Sigma_\theta^{-1}\dot{m}(\tau,\theta_0)'$$
$$+\dot{m}(v,\theta_0)'\Sigma_\theta^{-1}\left(\int \bar{l}(u,\theta_0)' H(du)\bar{l}(u,\theta_0)\right)\Sigma_\theta^{-1}\dot{m}(\tau,\theta_0).$$

Note that $dH(u) = \left(\operatorname{diag}(p(u_2)) - p(u_2)p(u_2)'\right) dF_u(u)$ such that

$$\begin{aligned}
\int \phi(u,\tau)\,dH(u)\bar{l}(u,\theta_0) &= \int \phi(u,\tau)\,dH(u)\left(\operatorname{diag}(p(u_2)) - p(u_2)p(u_2)'\right)^{-1}\frac{\partial p(u_2,\theta)}{\partial\theta'} \\
&= \int \phi(u,\tau)\frac{\partial p(u_2,\theta_0)}{\partial\theta} dF_u(u) = \dot{m}(\tau,\theta_0)
\end{aligned}$$

and

$$\begin{aligned}
&\int \bar{l}(u,\theta_0)' dH(u)\bar{l}(u,\theta_0) \\
= &\int \frac{\partial p'(u_2,\theta)}{\partial\theta}\left(\operatorname{diag}(p(u_2)) - p(u_2)p(u_2)'\right)^{-1}\frac{\partial p(u_2,\theta)}{\partial\theta'} dF_u(u) = \Sigma_\theta
\end{aligned}$$

such that $E\left[\hat{V}(v)\hat{V}(\tau)'\right] = H(v \wedge \tau) - \dot{m}(v,\theta_0)'\Sigma_\theta^{-1}\dot{m}(\tau,\theta_0)$ as required.

We now verify that the transformation T has the required properties. Note that

$$\begin{aligned}\left\langle \phi(.,v)', \bar{l}(.,\theta)\right\rangle &= \int \phi\left(u,v\right) dH(u) \left(\operatorname{diag}\left(p(u_2)\right) - p(u_2)p\left(u_2\right)'\right)^{-1} \frac{\partial p(u_2,\theta)}{\partial \theta'} \\ &= \dot{m}(v,\theta_0)\end{aligned}$$

such that $\hat{V}\left(v\right) = V(\phi\left(.,v\right)) - \left\langle \phi\left(.,\tau\right)', \bar{l}(.,\theta)\right\rangle C_{-\infty}^{-1} V(\bar{l}(v,\theta))$.

In order to establish that $T\hat{V}(v) = \hat{V}\left(v\right) - \int \left\langle \phi\left(.,v\right)', d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} \hat{V}(\pi_\lambda^{\perp} \bar{l}(.,\theta))$ has covariance function $\Gamma(v,\tau)$ we first consider $E\left(TV(v)\right)^2$ where

$$\begin{aligned}& E\left(V\left(v\right) - \int \left\langle \phi\left(.,v\right)', d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} \int \pi_\lambda^{\perp} \bar{l}(\vartheta,\theta)' db(u)\right)^2 \\ =\ & \Gamma(v,v) - 2\int \left\langle \phi\left(.,v\right)', d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} \left\langle \pi_\lambda^{\perp} \bar{l}(.,\theta), \phi\left(.,v\right)'\right\rangle \\ & + \int\int \left\langle \phi\left(.,v\right), d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} \left(\int \pi_\lambda^{\perp} \bar{l}(u,\theta)' dH(u) \pi_\mu^{\perp} \bar{l}(u,\theta)\right) C_\mu^{-1} \left\langle d\pi_\mu \bar{l}(.,\theta), \phi\left(.,v\right)'\right\rangle \\ =\ & \Gamma(v,v) - 2\int \left\langle \phi\left(.,v\right), d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} \left\langle \pi_\lambda^{\perp} \bar{l}(.,\theta), \phi\left(.,v\right)'\right\rangle \\ & + \int\int \left\langle \phi\left(.,v\right), d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} C_{\lambda\vee\mu} C_\mu^{-1} \left\langle d\pi_\mu \bar{l}(.,\theta), \phi\left(.,v\right)'\right\rangle.\end{aligned}$$

Note that $\left\langle \phi\left(.,v\right), d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} C_{\lambda\vee\mu} C_\mu^{-1} \left\langle d\pi_\mu \bar{l}(.,\theta), \phi\left(.,v\right)'\right\rangle$ is symmetric in λ and μ such that

$$\begin{aligned}& \int\int \left\langle \phi\left(.,v\right), d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} C_{\lambda\vee\mu} C_\mu^{-1} \left\langle \phi\left(.,v\right), d\pi_\mu \bar{l}(.,\theta)'\right\rangle \\ =\ & 2\int \left\langle \phi\left(.,v\right), d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} \int_\lambda^\infty \left\langle d\pi_\mu \bar{l}(.,\theta), \phi\left(.,v\right)'\right\rangle \\ =\ & 2\int \left\langle \phi\left(.,v\right), d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} \left\langle \pi_\lambda \bar{l}(.,\theta), \phi\left(.,v\right)'\right\rangle\end{aligned}$$

such that $E\left(V\left(v\right) - \int \left\langle \phi\left(.,v\right), d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} V(\pi_\lambda^{\perp} \bar{l}(.,\theta))\right)^2 = \Gamma\left(v,v\right)$. By the same arguments it follows that $E\left[TV(v)TV(\tau)\right] = \Gamma\left(v,\tau\right)$.

That the result then also holds for $T\hat{V}\left(v\right)$ follows from Khmaladze (1993, Theorem 3.9). ∎

Khmaladze (1993, Lemmas 3.2-3.4) shows that the argument need not be limited to all v such that $v \in \Upsilon_x$. As noted by Koul and Stute, however, once T is replaced by T_n convergence can only be shown on the subset $\pi_x v$ of $[-\infty,\infty]^k$ for some finite x due to the instability of the estimated matrix C_λ as $\lambda \to \infty$.

The next step is to analyze the transform T when applied to the empirical processes $V_n(v)$ and $\hat{V}_n(v)$ and in particular to show convergence to the limiting counterpart, $T\hat{V}(v)$.

Proposition 5 *Assume Conditions 3, 4, 5, 6, 7, 8 and 9 are satisfied. Fix $x < \infty$ arbitrary and define $\Upsilon_x = \left\{ v \in [-\infty, \infty]^k \, | v = \pi_x v \right\}$. Then,*

$$\sup_{v \in \Upsilon_x} \left| T\hat{V}_n(v) - TV_n(v) \right| = o_p(1)$$

and $TV_n(v) \Rightarrow TV(v)$ in $\mathfrak{D}[\Upsilon_x]$ where $\Rightarrow$ denotes weak convergence.

Proof of Proposition 5. By Theorem 3 we have uniformly on $[-\infty, \infty]^k$ that $\hat{V}_n(v) - V_n(v) = \dot{m}(v, \theta_0) n^{-1/2} \sum_{t=1}^n l(D_t, z_t, \theta_0) + o_p(1)$. Thus consider the difference

$$\begin{aligned}
& T\hat{V}_n - TV_n \qquad (27) \\
= \; & -\dot{m}(v, \theta_0) n^{-1/2} \sum_{t=1}^n l(D_t, z_t, \theta_0) \\
& - \int \left\langle \phi(., v)', d\pi_\lambda \bar{l}(., \theta_0) \right\rangle C_\lambda^{-1} \left(\hat{V}_n \left(\pi_\lambda^\perp \bar{l}(., \theta_0)' \right) - V_n \left(\pi_\lambda^\perp \bar{l}(., \theta_0)' \right) \right) + o_p(1)
\end{aligned}$$

where $\hat{H}_n$ and H_n are defined in Appendix B.1 for $\|\theta_n - \theta_0\| \leq \left\| \hat{\theta} - \theta_0 \right\|$ it follows by the mean value theorem that

$$\begin{aligned}
& \hat{V}_n \left(\pi_\lambda^\perp \bar{l}(., \theta_0)' \right) - V_n \left(\pi_\lambda^\perp \bar{l}(., \theta_0)' \right) \\
= \; & n^{-1/2} \sum_{t=1}^n \mathbf{1}\{U_t \notin A_\lambda\} \bar{l}(z_t, \theta_0)' \left(p(z_t, \theta_0) - p(z_t, \hat{\theta}) \right) \\
= \; & n^{-1/2} \sum_{t=1}^n \mathbf{1}\{U_t \notin A_\lambda\} \bar{l}(z_t, \theta_0)' \left(\frac{\partial p(z_t, \theta_n)}{\partial \theta'} - \frac{\partial p(z_t, \theta_0)}{\partial \theta'} \right) \left(\hat{\theta} - \theta_0 \right) \\
& + n^{-1/2} \sum_{t=1}^n \mathbf{1}\{U_t \notin A_\lambda\} \bar{l}(z_t, \theta_0)' \frac{\partial p(z_t, \theta_0)}{\partial \theta'} \left(\hat{\theta} - \theta_0 \right) \\
: \; & = R_1(\lambda) + R_2(\lambda).
\end{aligned}$$

Let $\dot{m}(\theta) = E\left[\frac{\partial p(z_t, \theta)}{\partial \theta'} \right]$ and $\dot{m}(z_t, \theta) = \frac{\partial p(z_t, \theta)}{\partial \theta} - \dot{m}(\theta)$. First consider

$$\begin{aligned}
\sup_\lambda \|R_1(\lambda)\| \leq \; & n^{1/2} \left\| \hat{\theta} - \theta_0 \right\| n^{-1} \sum_{t=1}^n \left\| \bar{l}(z_t, \theta_0) \right\| \left\| \frac{\partial p(z_t, \theta_n)}{\partial \theta'} - \frac{\partial p(z_t, \theta_0)}{\partial \theta'} \right\| \\
\leq \; & n^{1/2} \left\| \hat{\theta} - \theta_0 \right\| n^{-1} \sum_{t=1}^n \left\| \bar{l}(z_t, \theta_0) \right\| \|\dot{m}(z_t, \theta_n) - \dot{m}(z_t, \theta_0)\| \\
& + n^{1/2} \left\| \hat{\theta} - \theta_0 \right\| n^{-1} \sum_{t=1}^n \left\| \bar{l}(z_t, \theta_0) \right\| \|\dot{m}(\theta_n) - \dot{m}(\theta_0)\| \\
\leq \; & n^{1/2} \left\| \hat{\theta} - \theta_0 \right\| \left(n^{-1} \sum_{t=1}^n \left\| \bar{l}(z_t, \theta_0) \right\|^2 \right)^{1/2} \left(n^{-1} \sum_{t=1}^n \|\dot{m}(z_t, \theta_n) - \dot{m}(z_t, \theta_0)\|^2 \right)^{1/2}
\end{aligned}$$

where the third inequality follows from Hölder's inequality. Since $\|\theta_n - \theta_0\| = o_p(1)$ it follows from the continuous mapping theorem that $\|\dot{m}(\theta_n) - \dot{m}(\theta_0)\| = o_p(1)$. Together with the fact that $E\left\|\bar{l}(z_t,\theta_0)\right\| < \infty$ and Lemma 2.1 of Arcones and Yu (1994) this implies that

$$n^{1/2}\left\|\hat{\theta} - \theta_0\right\| n^{-1}\sum_{t=1}^{n}\left\|\bar{l}(z_t,\theta_0)\right\| \left\|\dot{m}(\theta_n) - \dot{m}(\theta_0)\right\| = o_p(1).$$

By Condition 7 it follows that

$$\|\dot{m}(z_t,\theta_n) - \dot{m}(z_t,\theta_0)\|^2 \le k\,|B(z_t)|^2\,\|\theta_n - \theta_0\|^{2\alpha}$$

for some $\alpha > 0$ such that

$$n^{-1}\sum_{t=1}^{n}\|\dot{m}(z_t,\theta_n) - \dot{m}(z_t,\theta_0)\|^2 \le k\,\|\theta_n - \theta_0\|^{2\alpha}\, n^{-1}\sum_{t=1}^{n}|B(z_t)|^2 = o_p(1).$$

This establishes $\sup_\lambda \|R_1(\lambda)\| = o_p(1)$ such that uniformly on Υ_x

$$\left\|\int \left\langle \phi(.,v)', d\pi_\lambda \bar{l}(.,\theta_0)\right\rangle C_\lambda^{-1} R_1(\lambda)\right\| \le \sup_\lambda \|R_1(\lambda)\| \sup_{\lambda:\pi_\lambda\in\Upsilon_x}\|C_\lambda\|^{-1}\int \left\|\left\langle \phi(.,v)', d\pi_\lambda \bar{l}(.,\theta_0)\right\rangle\right\| = o_p(1).$$

Next consider $R_2(\lambda) - \left(\int \pi_\lambda^\perp \bar{l}(\vartheta,\theta_0)'\frac{\partial p(u_2,\theta_0)}{\partial\theta'}dF_u(u)\right) n^{1/2}\left(\hat{\theta} - \theta_0\right)$. Note that

$$E\left[\mathbf{1}\{U_t \notin A_\lambda\}\,\bar{l}(z_t,\theta_0)\frac{\partial p(z_t,\theta_0)}{\partial\theta'}\right] = \int \pi_\lambda^\perp \bar{l}(u,\theta_0)'\frac{\partial p(u_2,\theta_0)}{\partial\theta'}dF_u(u)$$

and

$$\begin{aligned}
& \sup_\lambda \left\|\mathbf{1}\{U_t \notin A_\lambda\}\,\bar{l}(z_t,\theta_0)'\frac{\partial p(z_t,\theta_0)}{\partial\theta'}\right\| \\
\le\ & \left\|\bar{l}(z_t,\theta_0)'\frac{\partial p(z_t,\theta_0)}{\partial\theta'}\right\| \\
=\ & \left\|\frac{\partial p'(z_t,\theta_0)}{\partial\theta}\left(\operatorname{diag}(p(u_2)) - p(u_2)p(u_2)'\right)^{-1}\frac{\partial p(z_t,\theta_0)}{\partial\theta'}\right\| \\
\le\ & \left\|\frac{\partial p'(z_t,\theta_0)}{\partial\theta}\right\|^2\left\|\left(\operatorname{diag}(p(u_2)) - p(u_2)p(u_2)'\right)^{-1/2}\right\|^2 \\
\le\ & \left(\sup_{u_2}\left(\mathbf{1}_{\mathcal{M}}'\left(\operatorname{diag}(p(u_2)) - p(u_2)p(u_2)'\right)^{-1}\mathbf{1}_{\mathcal{M}}\right)\right)\sum_{i=1}^{\mathcal{M}}\sum_{j=1}^{d}\left(\frac{\partial p_i(z_t,\theta_0)}{\partial\theta_j}\right)^2
\end{aligned}$$

where $\left(\sup_{u_2}\left(\mathbf{1}_{\mathcal{M}}'\left(\operatorname{diag}(p(u_2)) - p(u_2)p(u_2)'\right)^{-1}\mathbf{1}_{\mathcal{M}}\right)\right)$ is bounded by Condition 6 and $E\left[\left(\frac{\partial p_i(z_t,\theta_0)}{\partial\theta_j}\right)^2\right]$ is bounded by Condition 7. This shows that $(1-\mathbf{1}\{(y_t,z_t)\in A_\lambda\})\,\bar{l}(z_t,\theta_0)\frac{\partial p(z_t,\theta_0)}{\partial\theta'}$ is a Euclidean class with integrable envelope $\left\|\bar{l}(z_t,\theta_0)\frac{\partial p(z_t,\theta_0)}{\partial\theta'}\right\|$ such that by Lemma 2.1 of Arcones and Yu it follows that

$$\sup_\lambda\left\|R_2(\lambda) - \left(\int \pi_\lambda^\perp \bar{l}(\vartheta,\theta_0)'\frac{\partial p(u_2,\theta_0)}{\partial\theta'}dF_u(u)\right) n^{-1/2}\left(\hat{\theta} - \theta_0\right)\right\| = o_p(1).$$

It then follows that uniformly on Υ_x

$$\int \left\langle \phi\left(.,v\right), d\pi_\lambda \bar{l}(.,\theta_0)\right\rangle C_\lambda^{-1} \left[R_2\left(\lambda\right) - \int \pi_\lambda^\perp \bar{l}(\vartheta,\theta_0)' \frac{\partial p(u_2,\theta_0)}{\partial\theta'} dF_u\left(u\right) n^{-1/2}\left(\hat{\theta}-\theta_0\right)\right] = o_p\left(1\right).$$

Now note that $\int \pi_\lambda^\perp \bar{l}(\vartheta,\theta_0)' \frac{\partial p(u_2,\theta_0)}{\partial\theta'} dF_u\left(u\right) = C_\lambda$ such that

$$\begin{aligned}
& \int \left\langle \phi\left(.,v\right)', d\pi_\lambda \bar{l}(.,\theta_0)\right\rangle C_\lambda^{-1} \int \pi_\lambda^\perp \bar{l}(\vartheta,\theta_0)' \frac{\partial p(u_2,\theta_0)}{\partial\theta'} dF_u\left(u\right) n^{-1/2}\left(\hat{\theta}-\theta_0\right) \\
= & \int \left\langle \phi\left(.,v\right)', d\pi_\lambda \bar{l}(.,\theta_0)\right\rangle n^{-1/2}\left(\hat{\theta}-\theta_0\right) \\
= & \dot{m}(v,\theta_0) n^{-1/2}\left(\hat{\theta}-\theta_0\right).
\end{aligned}$$

Substituting back in 27 then shows that $\sup_{v\in\Upsilon_x}\left|T\hat{V}_n\left(v\right) - TV_n\left(v\right)\right| = o_p(1)$.

For the second part of the proposition consider

$$TV_n\left(v\right) = V_n\left(v\right) - \int \left\langle \phi\left(.,v\right)', d\pi_\lambda \bar{l}(.,\theta_0)\right\rangle C_\lambda^{-1} n^{-1/2} \sum_{t=1}^n \mathbf{1}\left\{U_t \notin A_\lambda\right\} \bar{l}(z_t,\theta_0)' \left(\mathcal{D}_t - p(z_t,\theta_0)\right).$$

Under H_0 it follows that

$$\begin{aligned}
& E\left[\mathbf{1}\left\{U_t \notin A_\lambda\right\} \bar{l}(z_t,\theta_0)\left(\mathcal{D}_t - p(z_t,\theta_0)\right)|z_t\right] \\
= & E\left[\left(\mathcal{D}_t - p(z_t,\theta_0)\right)|z_t\right] E\left[\mathbf{1}\left\{U_t \notin A_\lambda\right\}|z_t\right] \bar{l}(z_t,\theta_0) = 0
\end{aligned}$$

such that $V_n\left(v\right)$ is a martingale. The finite dimensional distributions can therefore be obtained from a martingale difference CLT. Let

$$g(y_t,z_t,v) = \int \left\langle \phi\left(.,v\right), d\pi_\lambda \bar{l}(.,\theta_0)\right\rangle C_\lambda^{-1} \mathbf{1}\left\{U_t \notin A_\lambda\right\} \bar{l}(z_t,\theta_0)'$$

such that $TV_n(v) = n^{-1/2}\sum_{t=1}^n \left(\phi\left(U_t,v\right) - g(y_t,z_t,v)\right)\left(D_t - p(z_t,\theta_0)\right)$. Then let

$$\begin{aligned}
Y_{1t}\left(v\right) &= \phi\left(U_t,v\right)\left(D_t - p(z_t,\theta_0)\right), \\
Y_{2t}\left(v\right) &= g(y_t,z_t,v)\left(D_t - p(z_t,\theta_0)\right),
\end{aligned}$$

$Y_t\left(v\right) = Y_{1t}\left(v\right) - Y_{2t}\left(v\right)$ and $Y_{nt}\left(v\right) = n^{-1/2}Y_t\left(v\right)$. It follows that

$$E\left[Y_{1t}\left(v\right)Y_{1t}\left(v\right)'\right] = \Gamma(v,v),$$

$$\begin{aligned}
E\left[Y_{2t}\left(v\right)Y_{2t}\left(v\right)'\right] &= \int\int \{\left\langle \phi\left(.,v\right), d\pi_\lambda \bar{l}(.,\theta_0)\right\rangle C_\lambda^{-1} \\
& \times E\left[E\left[\mathbf{1}\left\{U_t \notin A_\lambda\right\}\mathbf{1}\left\{U_t \notin A_\mu\right\}|z_t\right] \frac{\partial p_t\left(z_t,\theta_0\right)}{\partial\theta'} \left(\operatorname{diag}\left(p(z_t)\right) - p(z_t)p\left(z_t\right)'\right)^{-1} \frac{\partial p_t\left(z_t,\theta_0\right)}{\partial\theta}\right] \\
& \times C_\mu^{-1}\left\langle \phi\left(.,v\right), d\pi_\mu \bar{l}(.,\theta_0)'\right\rangle\} \\
&= \int\int \left\langle \phi\left(.,v\right), d\pi_\lambda \bar{l}(.,\theta_0)\right\rangle C_\lambda^{-1} C_{\mu\vee\lambda} C_\mu^{-1} \left\langle \phi\left(.,v\right), d\pi_\mu \bar{l}(.,\theta_0)'\right\rangle \\
&= 2\int \left\langle \phi\left(.,v\right), d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} \left\langle \phi\left(.,v\right), \pi_\lambda^\perp \bar{l}(.,\theta)\right\rangle,
\end{aligned}$$

and

$$
\begin{aligned}
E\left[Y_{1t}\left(v\right)Y_{2t}\left(v\right)\right] &= \int\left\langle\phi\left(.,v\right),d\pi_{\lambda}\bar{l}(.,\theta_{0})\right\rangle C_{\lambda}^{-1}E\left[E\left[\mathbf{1}\left\{U_{t}\notin A_{\lambda}\right\}\phi\left(U_{t},v\right)|z_{t}\right]\frac{\partial p_{t}\left(z_{t},\theta_{0}\right)}{\partial\theta'}\right] \\
&= \int\left\langle\phi\left(.,v\right),d\pi_{\lambda}\bar{l}(.,\theta)\right\rangle C_{\lambda}^{-1}\left\langle\phi\left(U_{t},v\right),\pi_{\lambda}^{\perp}\bar{l}(.,\theta)\right\rangle
\end{aligned}
$$

which shows that $E\left[Y_{t}\left(v\right)Y_{t}\left(v\right)'\right]=\Gamma(v,v)$. Also, $E\left[Y_{1t}\left(v\right)Y_{1t}\left(\tau\right)'\right]=\Gamma\left(v,\tau\right)$,

$$
E\left[Y_{2t}\left(v\right)Y_{2t}\left(\tau\right)'\right]=2\int\left\langle\phi\left(.,v\right),d\pi_{\lambda}\bar{l}(.,\theta)\right\rangle C_{\lambda}^{-1}\left\langle\phi\left(.,\tau\right),\pi_{\lambda}^{\perp}\bar{l}(.,\theta)\right\rangle
$$

and

$$
E\left[Y_{1t}\left(v\right)Y_{2t}\left(\tau\right)'\right]=\int\left\langle\mathbf{1}\left(.\leq v\right),d\pi_{\lambda}\bar{l}(.,\theta)\right\rangle C_{\lambda}^{-1}\left\langle\mathbf{1}\left(.\leq\tau\right),\pi_{\lambda}^{\perp}\bar{l}(.,\theta)\right\rangle
$$

such that $E\left[Y_{t}\left(v\right)Y_{t}\left(\tau\right)'\right]=\Gamma\left(v,\tau\right)$. It also follows that $E\left\|Y_{t}\right\|^{2}<\infty$ such that the conditional Lindeberg condition of the CLT is satisfied. We conclude that the finite dimensional distributions of $TV_{n}(v)$ converge to a Gaussian limit with mean zero and covariance function $\Gamma(v,\tau)$. For weak convergence in the function space note that

$$
\begin{aligned}
\|g(y_{t},z_{t},v)\| &\leq \int\left\|\left\langle\phi\left(.,v\right),d\pi_{\lambda}\bar{l}(.,\theta_{0})\right\rangle C_{\lambda}^{-1}\bar{l}(z_{t},\theta_{0})'\right\| \\
&\leq \int\left\|\left\langle\phi\left(.,v\right),d\pi_{\lambda}\bar{l}(.,\theta_{0})\right\rangle C_{\lambda}^{-1}\right\|\left\|\bar{l}(z_{t},\theta_{0})\right\|
\end{aligned}
$$

where $\int\left\|\left\langle\phi\left(.,v\right),d\pi_{\lambda}\bar{l}(.,\theta_{0})\right\rangle C_{\lambda}^{-1}\right\|$ is uniformly bounded on Υ_{x} and $\left\|\bar{l}(z_{t},\theta_{0})\right\|^{2}=\sum_{j=1}^{\mathcal{M}}\sum_{i=1}^{d}\left|\bar{l}_{i,j}(z_{t},\theta_{0})\right|^{2}$ such that by the Hölder inequality

$$
\left|\bar{l}_{i,j}(z_{t},\theta_{0})\right|^{2+\delta}\leq(\mathcal{M}+1)^{1+\delta/2}\left(\left|\frac{\partial p_{i}(z_{t},\theta_{0})/\partial\theta_{j}}{p_{i}\left(z_{t},\theta_{0}\right)}\right|^{2+\delta}+\frac{\sum_{j=1}^{\mathcal{M}}\left|\partial p_{j}(z_{t},\theta_{0})/\partial\theta_{j}\right|^{2+\delta}}{\left|1-\sum_{j=1}^{\mathcal{M}}p_{j}(z_{t},\theta_{0})\right|^{2+\delta}}\right).
$$

By the Cauchy Schwartz inequality it then follows that

$$
\begin{aligned}
E\left|\bar{l}_{i,j}(z_{t},\theta_{0})\right|^{2+\delta} &\leq (\mathcal{M}+1)^{1+\delta/2}\left(E\left[\left|\partial p_{i}(z_{t},\theta_{0})/\partial\theta_{j}\right|^{4+2\delta}\right]\right)^{1/2}\left(E\left[p_{i}\left(z_{t},\theta_{0}\right)^{-(4+2\delta)}\right]\right)^{1/2} \\
&\quad+(\mathcal{M}+1)^{1+\delta/2}\sum_{j=1}^{\mathcal{M}}\left(E\left[\left|\partial p_{j}(z_{t},\theta_{0})/\partial\theta_{j}\right|^{4+2\delta}\right]\right)^{1/2}\left(E\left[\left|1-\sum_{j=1}^{\mathcal{M}}p_{j}(z_{t},\theta_{0})\right|^{-(4+2\delta)}\right]\right)^{1/2} \\
&< \infty
\end{aligned}
$$

which is bounded for some δ by Condition 7. This shows that $g(y_{t},z_{t},v)$ is a Euclidean class of functions and by Lemma 2.14 of Pakes and Pollard it follows that $Y_{t}(v)$ is a Euclidean class of functions. Lemma 2.1 of Arcones and Yu then can be used to establish weak convergence on $\mathfrak{D}\left[\Upsilon_{x}\right]$. ■

Our main formal result is established next.

Theorem 6 *Assume Conditions 3, 4, 5, 6,7, 8 and 9 are satisfied. Fix $x < \infty$ arbitrary and define $\Upsilon_x = \left\{ v \in [-\infty, \infty]^k \,|v = \pi_x v \right\}$. Then, for T_n defined in (12),*

$$\sup_{v \in \Upsilon_x} \left| T_n \hat{V}_n(v) - TV_n(v) \right| = o_p(1).$$

Proof of Theorem 6. We start by considering $\hat{C}_\lambda - C_\lambda$. Let

$$C_\lambda(\theta) = E\left[\mathbf{1}\{U_t \notin A_\lambda\} \bar{l}(z_t, \theta)' \frac{\partial p(z_t, \theta)}{\partial \theta'} \right]$$

such that $C_\lambda = C_\lambda(\theta_0)$ and

$$\begin{aligned} \hat{C}_\lambda - C_\lambda &= n^{-1} \sum_{t=1}^n \mathbf{1}\{U_t \notin A_\lambda\} \bar{l}(z_t, \hat{\theta})' \frac{\partial p(z_t, \hat{\theta})}{\partial \theta'} - C_\lambda(\theta_0) \\ &= n^{-1} \sum_{t=1}^n \mathbf{1}\{U_t \notin A_\lambda\} \bar{l}(z_t, \hat{\theta})' \frac{\partial p(z_t, \hat{\theta})}{\partial \theta'} - C_\lambda\left(\hat{\theta}\right) + C_\lambda\left(\hat{\theta}\right) - C_\lambda(\theta_0). \end{aligned}$$

Note that $C_\lambda(\theta) = \int (1 - \mathbf{1}(u \in A_\lambda)) \bar{l}(u, \theta)' H(du) \bar{l}(u, \theta)$ such that for any λ, θ it follows that

$$\begin{aligned} \left\| C_{\lambda'}(\theta') - C_\lambda(\theta) \right\| \leq& \left\| \int (\mathbf{1}(u \in A_{\lambda'}) - \mathbf{1}(u \in A_\lambda)) \bar{l}(u, \theta')' dH(u) \bar{l}(u, \theta') \right\| \\ &+ \left\| \int \mathbf{1}(u \in A_\lambda) \left(\bar{l}(u, \theta')' dH(u) \bar{l}(u, \theta') - \bar{l}(u, \theta)' dH(u) \bar{l}(u, \theta)' \right) \right\| \end{aligned}$$

where $|\mathbf{1}(u \in A_{\lambda'}) - \mathbf{1}(u \in A_\lambda)| \leq \mathbf{1}\left(u \in A_{\max(\lambda,\lambda')} \backslash A_{\min(\lambda,\lambda')}\right) \to 0$ as $\lambda' \to \lambda$ by Condition 9. Continuity of $\bar{l}(u, \theta)' \bar{l}(u, \theta)$ and integrability of the envelope function $\left\| \bar{l}(u, \theta_0) \right\|^2$ then establish uniform continuity of $C_\lambda(\theta)$ on $\Upsilon_x \times N(\theta_0)$ by use of the dominated convergence theorem. By continuity of $C_\lambda(\theta)$ and the continuous mapping theorem it now follows that $\left\| C_\lambda\left(\hat{\theta}\right) - C_\lambda(\theta_0) \right\| = o_p(1)$ uniformly on $\Upsilon_x \times N(\theta_0)$. Let $v_n(\theta, \lambda) = n^{-1} \sum_{t=1}^n \mathbf{1}\{U_t \notin A_\lambda\} \bar{l}(z_t, \theta)' \frac{\partial p(z_t, \theta)}{\partial \theta'} - C_\lambda(\theta)$. We note that

$$\left\| \mathbf{1}\{U_t \notin A_\lambda\} \bar{l}(z_t, \theta)' \frac{\partial p(z_t, \theta)}{\partial \theta'} \right\| \leq 2 \left\| \bar{l}(z_t, \theta) \right\|^2 \left\| \operatorname{diag}(p(z_t)) - p(z_t) p(z_t)' \right\| \leq 2\mathcal{M} \left\| \bar{l}(z_t, \theta) \right\|^2$$

where $\bar{l}_{i,j}(z_t, \theta)$ has the integrable Envelope $B(z_t) \left(2\sqrt{d} \sup_{N(\theta_0)} \left\| \theta - \theta' \right\| \right)^\alpha + \left| \bar{l}_{i,j}(z_t, \theta_0) \right|$ on $N(\theta_0)$ by Condition 7. By Condition 9 the functions $\mathbf{1}\{(y_t, z_t) \in A_\lambda\}$ form a Euclidean class. It now follows from Lemma 2.1 of Arcones and Yu (1994) that, because $n^{1/2} v_n(\theta, \lambda)$ converges weakly to a Gaussian limit, a tightness condition must hold, i.e. for any $\varepsilon, \eta > 0$, $\exists \delta > 0$ such that

$$\limsup_n \Pr\left(\sup_{\lambda, \theta \in \Upsilon_x \times N(\theta_0)} \sup_{\lambda', \theta' : d((\lambda, \theta), (\lambda', \theta')) < \delta} \left\| v_n(\theta', \lambda') - v_n(\theta, \lambda) \right\| > \varepsilon \right) < \eta. \tag{28}$$

Property 28 together with the boundedness of the space $\Upsilon_x \times N(\theta_0)$ now implies by a conventional approximation argument, that

$$\sup_{\lambda,\theta\in\Upsilon_x\times N(\theta_0)} \|v_n(\theta,\lambda)\| = o_p(1).$$

It now follows that

$$\Pr\left(\left\|\hat{C}_\lambda - C_\lambda\left(\hat{\theta}\right)\right\| > \varepsilon\right) \leq \Pr\left(\sup_{\lambda,\theta\in\Upsilon_x\times N(\theta_0)} \|v_n(\theta,\lambda)\| > \varepsilon\right) + \Pr\left(\hat{\theta}\notin N(\theta_0)\right) \xrightarrow{p} 0 \tag{29}$$

such that $\sup_{\lambda\in\Upsilon_x}\left\|\hat{C}_\lambda - C_\lambda\right\| = o_p(1)$.

Then

$$\begin{aligned} T_n\hat{V}_n(v) - TV_n(v) &= -\dot{m}(v,\theta_0)n^{-1/2}\sum_{t=1}^{n} l(D_t, z_t, \theta_0) + o_p(1) \\ &\quad - \int d\left(\int \phi(u,v)\, d\hat{H}_n(u)\pi_\lambda \bar{l}(.,\hat{\theta})\right)\hat{C}_\lambda^{-1}\hat{V}_n(\pi_\lambda^\perp \bar{l}(.,\hat{\theta})') \\ &\quad + \int \left\langle \phi(.,v), d\pi_\lambda \bar{l}(.,\theta_0)\right\rangle C_\lambda^{-1}V_n(\pi_\lambda^\perp \bar{l}(.,\theta_0)'). \end{aligned}$$

From before we have

$$\begin{aligned} &\int d\left(\int \phi(u,v)\, d\hat{H}_n(u)\pi_\lambda \bar{l}(.,\hat{\theta})\right)\hat{C}_\lambda^{-1}\hat{V}_n(\pi_\lambda^\perp \bar{l}(.,\hat{\theta})') \\ = &\int d\left(\int \phi(u,v)\, d\hat{H}_n(u)\pi_\lambda \bar{l}(.,\hat{\theta})\right)\left(\hat{C}_\lambda^{-1} - C_\lambda^{-1}\right)\hat{V}_n(\pi_\lambda^\perp \bar{l}(.,\hat{\theta})') \\ &+ \int d\left(\int \phi(u,v)\, d\hat{H}_n(u)\pi_\lambda \bar{l}(.,\hat{\theta})\right) C_\lambda^{-1}\hat{V}_n(\pi_\lambda^\perp \bar{l}(.,\hat{\theta})') \end{aligned}$$

where

$$\begin{aligned} &\left\|\int d\left(\int \phi(u,v)\, d\hat{H}_n(u)\pi_\lambda \bar{l}(.,\hat{\theta})\right)\left(\hat{C}_\lambda^{-1} - C_\lambda^{-1}\right)\hat{V}_n(\pi_\lambda^\perp \bar{l}(.,\hat{\theta}))\right\| \\ \leq &\sup_{\lambda\in[-\infty,x]}\left\|\hat{C}_\lambda^{-1} - C_\lambda^{-1}\right\| \int \left\|d\left(\int \phi(u,v)\, d\hat{H}_n(u)\pi_\lambda \bar{l}(.,\hat{\theta})\right)\right\| \left\|\hat{V}_n(\pi_\lambda^\perp \bar{l}(.,\hat{\theta}))\right\| = o_p(1) \end{aligned}$$

by 29. Next we consider

$$\begin{aligned}
\hat{V}_n(\pi_\lambda^\perp \bar{l}(.,\hat{\theta})') &= n^{-1/2}\sum_{t=1}^n \mathbf{1}\left\{U_t \notin A_\lambda\right\} \bar{l}(U_t,\hat{\theta})' \left(\mathcal{D}_t - p(z_t,\hat{\theta})\right) \\
&= n^{-1/2}\sum_{t=1}^n \mathbf{1}\left\{U_t \notin A_\lambda\right\} \bar{l}(U_t,\theta_0)' \left(\mathcal{D}_t - p(z_t,\theta_0)\right) \\
&\quad + \left[n^{-1/2}\sum_{t=1}^n \mathbf{1}\left\{U_t \notin A_\lambda\right\} \left((\mathcal{D}_t - p(z_t,\theta_0))' \otimes I_{\mathcal{M}}\right) \frac{\partial \operatorname{vec} \bar{l}(U_t,\theta_n)'}{\partial \theta'}\right] \left(\hat{\theta}-\theta_0\right) \\
&\quad - \left[n^{-1/2}\sum_{t=1}^n \mathbf{1}\left\{U_t \notin A_\lambda\right\} \bar{l}((y_t,z_t),\theta_0)' \frac{\partial p(z_t,\theta_n)}{\partial \theta'}\right] \left(\hat{\theta}-\theta_0\right) \\
&\quad - \left(\left(\hat{\theta}-\theta_0\right)' \otimes I_{\mathcal{M}}\right) \left[n^{-1/2}\sum_{t=1}^n \mathbf{1}\left\{U_t \notin A_\lambda\right\} \left(\frac{\partial p'(z_t,\theta_n)}{\partial \theta} \otimes I_{\mathcal{M}}\right) \frac{\partial \operatorname{vec} \bar{l}(U_t,\theta_n)}{\partial \theta'}\right] \left(\hat{\theta}-\theta_0\right) \\
&\equiv R_1(\lambda) + R_2(\lambda)\left(\hat{\theta}-\theta_0\right) + R_3(\lambda) n^{1/2}\left(\hat{\theta}-\theta_0\right) + n^{1/2}\left(\hat{\theta}-\theta_0\right)' R_4(\lambda)\left(\hat{\theta}-\theta_0\right)
\end{aligned}$$

where $\|\theta_n - \theta_0\| \leq \left\|\hat{\theta}-\theta\right\|$ and we have used the mean value theorem. Note that $R_1 = \int \pi_\lambda^\perp \bar{l}(\vartheta,\theta_0) dV_n(u)$,

$$\begin{aligned}
R_2(\lambda) &= n^{-1/2}\sum_{t=1}^n \mathbf{1}\left\{U_t \notin A_\lambda\right\} \left((\mathcal{D}_t - p(z_t,\theta_0))' \otimes I_{\mathcal{M}}\right) \frac{\partial \operatorname{vec} \bar{l}(U_t,\theta_0)'}{\partial \theta'} \\
&\quad + n^{-1/2}\sum_{t=1}^n \mathbf{1}\left\{U_t \notin A_\lambda\right\} \left((\mathcal{D}_t - p(z_t,\theta_0))' \otimes I_{\mathcal{M}}\right) \left(\frac{\partial \operatorname{vec} \bar{l}(U_t,\theta_n)'}{\partial \theta'} - \frac{\partial \operatorname{vec} \bar{l}(U_t,\theta_0)'}{\partial \theta'}\right) \\
&\equiv R_{21}(\lambda) + R_{22}(\lambda,\theta_n)
\end{aligned}$$

satisfies $ER_{21}(\lambda) = 0$ because

$$\begin{aligned}
& E\left[\mathbf{1}\left\{U_t \notin A_\lambda\right\} \left((\mathcal{D}_t - p(z_t,\theta_0))' \otimes I_{\mathcal{M}}\right) \frac{\partial \operatorname{vec} \bar{l}(U_t,\theta_0)'}{\partial \theta'} | z_t\right] \\
&= E\left[\left((\mathcal{D}_t - p(z_t,\theta_0))' \otimes I_{\mathcal{M}}\right) | z_t\right] E\left[\mathbf{1}\left\{U_t \notin A_\lambda\right\} \frac{\partial \operatorname{vec} \bar{l}(U_t,\theta_0)'}{\partial \theta'} | z_t\right] = 0
\end{aligned}$$

under H_0 such that finite dimensional convergence follows by the martingale difference CLT and uniform convergence follows from the fact that $\mathbf{1}\left\{U_t \notin A_\lambda\right\} \left((\mathcal{D}_t - p(z_t,\theta_0))' \otimes I_{\mathcal{M}}\right) \frac{\partial \operatorname{vec} \bar{l}(U_t,\theta_0)'}{\partial \theta'}$ is a Euclidean class of functions by Condition 9. It thus follows that $\sup_\lambda R_{21}(\lambda) = O_p(1)$ and $R_{21}(\lambda)\left(\hat{\theta}-\theta_0\right) = o_p(1)$ uniformly in λ. For the term $R_{22}(\lambda,\theta_n)$ we note that

$$E\left[\mathbf{1}\left\{U_t \notin A_\lambda\right\} \left((\mathcal{D}_t - p(z_t,\theta_0))' \otimes I_{\mathcal{M}}\right) \frac{\partial \operatorname{vec} \bar{l}(U_t,\theta)'}{\partial \theta'} | z_t\right] = 0$$

for any θ. By Lemma 2.1 of Arcones and Yu it thus follows that $R_{22}(\lambda,\theta)$ converges to a Gaussian limit process uniformly in λ and θ. Consequently, a tightness condition implied by this result can be used to show that $\limsup \Pr\left[\sup_{\theta:d(\theta,\theta_0)\leq\delta}\|R_{22}(\lambda,\theta)\| > \varepsilon\right] < \eta$ for all $\varepsilon, \eta > 0$ and some $\delta > 0$. Use root-n convergence of θ_n to conclude from this that $R_{22}(\lambda,\theta_n) = o_p(1)$. The terms involving θ_n in the remainder terms R_3 and R_4 containing θ_n can be handled in similar form and we therefore only consider the leading terms where θ_n is replaced by θ_0. For $R_4(\lambda)$ where

$$R_4(\lambda) = n^{-1}\sum_{t=1}^{n}\mathbf{1}\{U_t \notin A_\lambda\}\left(\frac{\partial p'(z_t,\theta_n)}{\partial\theta}\otimes I_{\mathcal{M}}\right)\frac{\partial \operatorname{vec}\bar{l}(U_t,\theta_n)}{\partial\theta'}$$

we note that $n^{1/2}(R_4(\lambda) - ER_4(\lambda))$ satisfies the conditions of Lemma 2.1 of Arcones and Yu (1994) such that it follows by similar arguments as before that $\sup_\lambda R_4(\lambda) = O_p(1)$. Then conclude that $n^{1/2}\left(\hat{\theta}-\theta_0\right)' R_4(\lambda)\left(\hat{\theta}-\theta_0\right) = o_p(1)$ uniformly in λ.

For $R_3(\lambda)$ note that

$$R_3(\lambda) = n^{-1}\sum_{t=1}^{n}\mathbf{1}\{U_t \notin A_\lambda\}\bar{l}(U_t,\theta_0)'\frac{\partial p(z_t,\theta_0)}{\partial\theta'}$$

uniformly converges to

$$ER_3(\lambda) = E\left[\mathbf{1}\{U_t \notin A_\lambda\}\bar{l}(U_t,\theta_0)\frac{\partial p(z_t,\theta_0)}{\partial\theta'}\right] = C_\lambda.$$

We have thus established that

$$\sup_\lambda\left\|\hat{V}_n(\pi_\lambda^\perp\bar{l}(.,\hat{\theta})) - V_n(\pi_\lambda^\perp\bar{l}(.,\theta_0)) - C_\lambda n^{1/2}\left(\hat{\theta}-\theta_0\right)\right\| = o_p(1).$$

Using this result we obtain

$$\begin{aligned}\int d\left(\int\phi(u,v)\,d\hat{H}_n(u)\pi_\lambda\bar{l}(u,\hat{\theta})\right) C_\lambda^{-1}\left(\hat{V}_n(\pi_\lambda^\perp\bar{l}(.,\hat{\theta})) - V_n(\pi_\lambda^\perp\bar{l}(.,\theta_0))\right)& \\ = \int d\left(\int\phi(u,v)\,d\hat{H}_n(u)\pi_\lambda\bar{l}(u,\hat{\theta})\right) n^{1/2}\left(\hat{\theta}-\theta_0\right) + o_p(1).&\end{aligned}$$

The leading term is then

$$\begin{aligned}\operatorname{vec}\int d\left(\int\phi(u,v)\,d\hat{H}_n(u)\pi_\lambda\bar{l}(u,\hat{\theta})\right) &= \operatorname{vec}\int d\left(\int\phi(u,v)\,dH_n(u)\pi_\lambda\bar{l}(u,\theta_0)\right) \\ &\quad+\int d\left(\int\phi(u,v)\,\pi_\lambda\frac{\partial\operatorname{vec}\partial p(u_2,\theta_n)/\partial\theta'}{\partial\theta'}d\hat{F}_u(u)\right)\left(\hat{\theta}-\theta_0\right)\end{aligned} \tag{30}$$

where $\hat{F}_u(u)$ is defined in (33) in Appendix B.1 and

$$\begin{aligned}
& \left\| \int d \int \phi\left(u,v\right) \pi_\lambda \frac{\partial \operatorname{vec} \partial p(u_2,\theta_n)/\partial\theta'}{\partial\theta'} d\hat{F}_u(u) \right\| \\
\leq \; & n^{-1} \sum_{t=1}^n \left\| \mathbf{1}\left\{U_t \leq v\right\} \mathbf{1}\left\{U_t \in A_\lambda\right\} \frac{\partial \operatorname{vec} \partial p(u_2,\theta_n)/\partial\theta'}{\partial\theta'} \right\| \\
\leq \; & n^{-1} \sum_{t=1}^n \left\| \frac{\partial \operatorname{vec} \partial p(u_2,\theta_0)/\partial\theta'}{\partial\theta'} \right\| \\
& + n^{-1} \sum_{t=1}^n \left\| \frac{\partial \operatorname{vec} \partial p(u_2,\theta_n)/\partial\theta'}{\partial\theta'} - \frac{\partial \operatorname{vec} \partial p(u_2,\theta_0)/\partial\theta'}{\partial\theta'} \right\| \\
\leq \; & n^{-1} \sum_{t=1}^n \left\| \frac{\partial \operatorname{vec} \partial p(u_2,\theta_0)/\partial\theta'}{\partial\theta'} \right\| + C \left\|\theta_n - \theta_0\right\|^\alpha n^{-1} \sum_{t=1}^n B(z_t) \\
= \; & O_p(1)
\end{aligned}$$

where C is a finite constant, the third inequality uses Condition 7 and the last equality follows from a standard law of large numbers for strong mixing sequences. The first term in 30 then is

$$\int d\left(\int \phi\left(u,v\right) dH_n(u) \pi_\lambda \bar{l}(u,\theta_0) \right) = n^{-1} \sum_{t=1}^n \phi\left(U_t,v\right) \mathbf{1}\left\{U_t \in A_\lambda\right\} \frac{\partial p(z_t,\theta_0)}{\partial\theta'}$$

where $E\left[\phi\left(U_t,v\right) \mathbf{1}\left\{U_t \in A_\lambda\right\} \frac{\partial p(z_t,\theta_0)}{\partial\theta}\right] = \dot{m}(\pi_\lambda v,\theta_0)$ for $v \in \Upsilon_x$. It thus follows again by a law or large numbers that $\int d\left(\int \phi\left(u,v\right) dH_n(u)\pi_\lambda \bar{l}(u,\theta_0)\right) = \dot{m}(v,\theta_0) + o_p\left(1\right)$ uniformly on Υ_x.

Finally we need to show that

$$\int \left(d\left(\int \phi\left(u,v\right) dH_n(u) \pi_\lambda \bar{l}(u,\theta_0) \right) - \left\langle \phi\left(.,v\right)', d\pi_\lambda \bar{l}(.,\theta_0) \right\rangle \right) C_\lambda^{-1} V_n(\pi_\lambda^\perp \bar{l}(u,\theta_0)) = o_p(1). \tag{31}$$

Let $g(z_t,\lambda,v) = \phi\left(U_t,v\right) \mathbf{1}\left\{U_t \in A_\lambda\right\} \frac{\partial p(z_t,\theta_0)}{\partial\theta}$. We first note that uniformly in λ on $[-\infty,x]$ and $v \in \Upsilon_x$,

$$\int \phi\left(.,v\right) \pi_\lambda dH_n(v) \bar{l}(.,\theta_0) - \left\langle \phi\left(.,v\right)', \pi_\lambda \bar{l}(.,\theta_0) \right\rangle = n^{-1} \sum_{t=1}^n g(z_t,\lambda,v) - E\left(g(z_t,\lambda,v)\right) \to 0 \text{ a.s.}$$

Weak convergence of $C_\lambda^{-1} V_n(\pi_\lambda^\perp \bar{l}(u,\theta_0))$ uniformly in λ on $[-\infty,x]$ can be established by the same methods as for $TV_n(v) \Rightarrow TV(v)$ in the second part of the proof of Proposition 5. We can thus proceed in the same way as Koul and Stute (1999, Lemma 4.2). Let $G_n(\lambda,v) = n^{-1}\sum_{t=1}^n g(z_t,\lambda,v)$, $G(\lambda,v) = E\left(g(z_t,\lambda,v)\right)$ and let $\zeta_n(\lambda) = C_\lambda^{-1} V_n(\pi_\lambda^\perp \bar{l}(u,\theta_0)')$. Then each component $\zeta_{ni}(\lambda)$ of the vector $\zeta_n(\lambda)$ is asymptotically tight by Prohorov's Theorem. In other words there exists a compact set $\mathbb{H}$ such that $\zeta_{ni}(\lambda) \in \mathbb{H}$ with probability no less than $1-\eta$ for any $\eta > 0$. Following the proof of Lemma 3.1 of Chang (1990) we choose step functions $a_1, a_2, ..., a_k \in \mathfrak{D}\left[-\infty,x\right]$ such that for any $\zeta \in \mathbb{H}$, $\sup\left|a_i - \zeta\right| < \varepsilon$ for some $i, 1 \leq i \leq k$. The right hand side of 31 can now be written as $\int_{-\infty}^{x} \zeta_n(\lambda)' (G_n(d\lambda) - G(d\lambda))$ such that for any $\delta > 0$

$$\begin{aligned}\Pr\left(\left\|\int_{-\infty}^{x} \zeta_n(\lambda)' (G_n(d\lambda) - G(d\lambda))\right\| > \eta\right) &\leq \Pr\left(\sup_{\zeta\in\mathbb{H}, v\in\Upsilon_x}\left\|\int_{-\infty}^{x} \zeta(\lambda)' (G_n(d\lambda, v) - G(d\lambda, v))\right\| > \delta\right) \\ &\quad + \Pr(\zeta_n \notin H).\end{aligned}$$

Since $\zeta \in \mathbb{H}$ it follows that

$$\sup_{\zeta\in\mathbb{H}, v\in\Upsilon_x}\left|\int_{-\infty}^{x} \zeta(\lambda)' (G_n(d\lambda, v) - G(d\lambda, v))\right| \leq \sup_{\zeta\in\mathbb{H}} \|\zeta(\lambda)\| \left(\sup_{v\in\Upsilon_x}\int_{-\infty}^{x} \|G(d\lambda, v)\| + \sup_{v\in\Upsilon_x}\int_{-\infty}^{x} \|G_n(d\lambda, v)\|\right)$$

where $\int_{-\infty}^{x} \|G(d\lambda, v)\| = \|G(x, v)\|$ and $\int_{-\infty}^{x} \|G_n(d\lambda, v)\| = \|G_n(x, v)\|$. Since $G(x, v) \to 0$ uniformly in v as $x \to -\infty$ and $G_n(\lambda, v)$ converges uniformly to $G(x, v)$ we can focus on a subset $[x_u, x] \subset [-\infty, x]$ where x_u is such that

$$\sup_{\zeta\in\mathbb{H}, v\in\Upsilon_x}\left\|\int_{-\infty}^{x_u} \zeta(\lambda)' (G_n(d\lambda, v) - G(d\lambda, v))\right\| < \delta$$

with probability tending to one. Now, for any component i, there exists a strictly increasing, continuous mapping κ of $[-\infty, x]$ onto itself, depending on ζ_i such that $\sup_{-\infty\leq\lambda\leq x} |\kappa(\lambda) - \lambda| < \varepsilon$ and $\sup_{-\infty\leq\lambda\leq x} |\zeta_i(\lambda) - a_i(\kappa(\lambda))| < \varepsilon$. Then for any element i, j of $\zeta(\lambda)' (G_n(d\lambda, v) - G(d\lambda, v))$

$$\begin{aligned}\left|\int_{x_u}^{x} \zeta_i(\lambda) (G_{ni,j}(d\lambda, v) - G_i(d\lambda, v))\right| &\leq \left|\int_{x_u}^{x} (\zeta_i(\lambda) - a_i(\kappa(\lambda))) (G_{ni,j}(d\lambda, v) - G_{i,j}(d\lambda, v))\right| \\ &\quad + \left|\int_{x_u}^{x} a_i(\kappa(\lambda)) (G_{ni,j}(d\lambda, v) - G_{i,j}(d\lambda, v))\right|\end{aligned}$$

which implies that for some N_0 and all $n > N_0$, $\left|\int_{-\infty}^{x} \zeta_i(\lambda) (G_{ni,j}(d\lambda, v) - G_{i,j}(d\lambda, v))\right| < 3\varepsilon$ uniformly on $H \times \Upsilon_x$ by the arguments of Chang (1994, p.396) which establishes 31. This now implies that $T_n\hat{V}_n(v) - TV_n(v) = o_p(1)$. ■

Theorem 6 together with Propositions 5 and 4 implies that $\hat{W}_n(v) - V_n(v) = o_p(1)$ uniformly in $v \in \Upsilon_x$. This in turn means that the limiting distribution of $\hat{W}_n(v)$ is a zero mean Gaussian process with covariance function $H(v, \tau)$. This distribution is not nuisance parameter free but can be computed conditional on the sample relatively easily as pointed out in Section 4.

Section 4.2 introduced the distribution free statistic $\hat{B}_{w,n}(w)$, defined as $\hat{B}_{w,n}(w) = \hat{W}_{w,n}\left(\phi(., w)h_w(.)^{-1/2}\right)$. By the arguments preceding Theorem 6, it follows that $\hat{B}_{w,n}(w) \Longrightarrow B_w(w)$ on $\Upsilon_{[0,1]}$. The only adjustments necessary are a restriction of $[-\infty, \infty]^k$ to $[0, 1]^k$. What remains to be shown is that

$$\sup_{v\in\Upsilon_{[0,1]}}\left|\hat{B}_{\hat{w},n}(w) - \hat{B}_{w,n}(w)\right| = o_p(1). \tag{32}$$

This is done in the next Theorem. We impose the following assumptions on the kernel function and density.

Condition 10 *The density $f_u(u)$ is continuously differentiable to some integral order $\omega \geq \max(2,k)$ on $\mathbb{R}^k$ with $\sup_{x\in\mathbb{R}^k} |D^\mu h(x)| < \infty$ for all $|\mu| \leq \omega$ where $\mu = (\mu_1, ..., \mu_k)$ is a vector of non-negative integers, $|\mu| = \sum_{j=1}^k \mu_j$, and $D^\mu f(x) = \partial^{|\mu|} h(x)/\partial x_1^{\mu_1} ....\partial x_k^{\mu_k}$ is the mixed partial derivative of order $|\mu|$. The kernel $K(.)$ 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 μ with $|\mu| \leq \omega$, $K(x) \to 0$ as $\|x\| \to \infty$ and $\sup_{x\in \mathbb{R}^k} (1 + \|x\|) |D^{e_i} K(x)| < \infty$ 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 $\int |\mathfrak{K}(r)| dr < \infty$ where $i = \sqrt{-1}$.*

Theorem 7 *Assume Conditions 3, 4, 5, 6,7, 8, 9 and 10 are satisfied. Fix $x < 1$ arbitrary and define $\Upsilon_{[0,1]} = \left\{ w \in [0,1]^k \,|w = \pi_x w \right\}$. Then,*

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

Proof of Theorem 7:. By Theorem 1 of Andrews (1995) it follows that

$$\sup_x \left| \hat{F}_k(x_k|x_{k-1}, ..., x_1) - F_k(x_k|x_{k-1}, ..., x_1) \right| = O_p(T^{-1/2} m_n^{-k}) + O_p(m_n^\omega).$$

By Pakes and Pollard (1989, Lemma 2.15) it follows that the composition of a function from a Euclidean class with envelope M and a measurable map with envelope M_1 forms another Euclidean class with envelope $M \circ M_1$. Since $F_k(x_k|x_{k-1}, ..., x_1)$ takes values in $[0,1]$ it clearly has an envelope M_1. It follows that $\hat{W}_{w,n}$ is a sample average over functions that belong to a Euclidean class plus remainder terms that vanish by similar arguments as before. It thus follows by the same arguments as before that for all $\varepsilon, \delta > 0$ there exists an $\eta > 0$ such that

$$\limsup_n \Pr \left(\sup_{\substack{w,w'\in\Upsilon_{[0,1]}, \|w-w'\|<\eta, \\ w_1,w_1'\in[0,1]^k, \|w_1-w_1'\|<\eta}} \left| \hat{B}_{w_1,n}(w) - \hat{B}_{w_1',n}(w') \right| > \varepsilon \right) < \delta.$$

It then follows that $\hat{B}_n(s) \Rightarrow B(s)$. ■

This result allows us to conduct inference using critical values that do not depend on nuisance parameters. Although these critical values must be calculated numerically, they are invariant to the sample distribution for a given design.

Theorem 8 *Assume Conditions 3, 4, 5, 6,7, 8, 9 and 10 are satisfied. Fix $x < 1$ arbitrary and define $\Upsilon_{[0,1]} = \left\{ w \in [0,1]^k \,|w = \pi_x w \right\}$. For $\hat{B}^*_{\hat{w},n}(w)$ defined in (20) it follows that $\hat{B}^*_{\hat{w},n}(w)$ converges on $\Upsilon_{[0,1]}$ to a Gaussian process $B_w(w)$.*

Proof. Following Chen and Fan (1999) we note that conditional on the data, $\hat{B}^{*}_{\hat{w},n}(w)$ is a Gaussian process with covariance function given by

$$\hat{\Gamma}_w(v,\tau) = n^{-1}\sum_{t=1}^{n}\left(m_{T,t}\left(v,\hat{\theta}\right) - \bar{m}_{n;T}\left(v,\hat{\theta}\right)\right)\left(m_{T,t}\left(\tau,\hat{\theta}\right) - \bar{m}_{n;T}\left(\tau,\hat{\theta}\right)\right)'.$$

By (32) and similar arguments as in the proof of Proposition 4 and Theorems 5 and 6 it follows that $\hat{\Gamma}_w(v,\tau)$ converges uniformly on $\Upsilon_{[0,1]}$ to the covariance function of $B_w(w)$, $\int \phi(u,v)\phi(u,\tau)du$. The result then follows in the same way as Theorem 5.2 of Chen and Fan (1999). ■

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 (14). We also present results for general sets A_λ. We start by defining the empirical distribution

$$\hat{F}_u(v) = n^{-1}\sum_{t=1}^{n}\{U_t \leq v\}, \tag{33}$$

and let

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

as well as

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

We now use the sets A_λ and projections π_λ as defined in Section 4.1. Let

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

such that

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

where

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

Finally, write

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

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

$$\hat{C}_\lambda = n^{-1}\sum_{t=1}^{n}\mathbf{1}\left\{y_{1t}>\lambda\right\}\bar{l}(z_t,\hat{\theta})\left(\operatorname{diag}\left(p(z_t,\hat{\theta})\right) - p(z_t,\hat{\theta})p\left(z_t,\hat{\theta}\right)'\right)\bar{l}(z_t,\hat{\theta})', \tag{34}$$

$$\hat{V}_n(\pi_\lambda^\perp \bar{l}(u,\hat{\theta})) = n^{-1/2}\sum_{t=1}^{n}\mathbf{1}\left\{y_{1t}>\lambda\right\}\bar{l}(U_t,\hat{\theta})'\left(\mathcal{D}_t - p(z_t,\hat{\theta})\right) \tag{35}$$

and

$$\int \phi(u,v)d\hat{H}_n(u)\pi_\lambda \bar{l}(u,\hat{\theta}) = n^{-1}\sum_{t=1}^{n}\mathbf{1}\left\{y_{1t}\le\lambda\right\}\phi\left\{U_t,v\right\}\frac{\partial p(z_t,\hat{\theta})}{\partial\theta'} \tag{36}$$

Combining 34, 35 and 36 then leads to the formulation 14.

B.2 Details for the Rosenblatt Transform

As before implementation requires replacement of θ 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

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

such that $\dot{m}(w,\theta)$ evaluated at the true parameter value θ_0 is

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

It therefore follows that $\hat{V}_{w,n}(v)$ can be approximated by $V_{w,n}(v) - \dot{m}_w(w,\theta_0)'n^{-1/2}\sum_{t=1}^{n} l\left(\mathcal{D}_t,z_t,\theta_0\right)$. 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 $h_w(.,\theta) = \left(\mathrm{diag}\left(p(\left[T_R^{-1}(.),\theta\right]_z)\right) - p(\left[T_R^{-1}(.),\theta\right]_z)p(\left[T_R^{-1}(.)\right]_z,\theta)'\right)$ and $h_w(.) := h_w(.,\theta_0)$.

We represent $\hat{V}_w$ in terms of V_w. Let $V_w(l_w(.,\theta_0)) = \int l_w(w,\theta_0)b_w(dv)$ 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

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

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

Let $\{A_{w,\lambda}\}$ be a family of measurable subsets of $[0,1]^k$, indexed by $\lambda\in[0,1]$ such that $A_{w,0}=\varnothing$, $A_{w,1}=[0,1]^k$, $\lambda\le\lambda' \Longrightarrow A_{w,\lambda}\subset A_{w,\lambda'}$ and $A_{w,\lambda'}\backslash A_{w,\lambda}\to\varnothing$ as $\lambda'\downarrow\lambda$. We then define the inner product $\langle f(.),g(.)\rangle_w := \int_{[0,1]^k} f(w)'dH_w(w)g(w)'$ where

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

and the matrix

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

and define the transform $T_wV_w(w)$ as before by

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

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 $W(v)$. In particular, using the notation $W_w(\phi(.,w)) = W_w(w)$ to emphasize the dependence of W on ϕ, it follows from the previous discussion that

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

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

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

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

C Model Definitions

- Models (a) to (f) fit an ordered Probit model to the change in the discretized intended federal funds rate ($\mathrm{d}Dff_t$).

  - 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, $\mathrm{gray0}_t$, $\mathrm{gray1}_t$, $\mathrm{gray2}_t$, igrym_t, $\mathrm{igry0}_t$, $\mathrm{igry1}_t$, $\mathrm{igry2}_t$, gradm_t, $\mathrm{grad0}_t$, $\mathrm{grad1}_t$, $\mathrm{grad2}_t$, igrdm_t, $\mathrm{igrd0}_t$, $\mathrm{igrd1}_t$, $\mathrm{igrd2}_t$, and our constructed unemployment innovation. For variable names, see the variable names appendix.
  - 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, $\mathrm{gray0}_t$, $\mathrm{gray1}_t$, $\mathrm{gray2}_t$, $\mathrm{igry0}_t$, $\mathrm{igry1}_t$, $\mathrm{igry2}_t$, $\mathrm{grad2}_t$, and our constructed unemployment innovation.
  - Specification (c) adds a quadratic term in the lagged intended federal funds rate change to the restricted model (b).
  - 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) to (h) use a Taylor specification in an ordered probit model for the the change in the discretized intended federal funds rate ($\mathrm{d}Dff_t$).

  - 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.
  - Specification (h) replaces dff_{t-2} with $(\mathrm{dff}_{t-1})^2$ in specification (g).
  - Specifications (i) and (j) are versions of (g) and (h) where covariates based on dff_t are replaced by covariates based on $dDff_t$.

D 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}	kth lag of GDP growth
inf_{t-k}	kth lag of inflation

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.

Table 1a: Parametric Sims-causality Tests for models using dff_{t-1}

Horizon		Model (a)	(b)	(c)	(g)	(h)
Quarter Lead	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 ***
Half Year Lead	1	0.25	0.17	0.73	0.60	1.45
	2	-1.30	-1.55	-0.40	-1.40	-0.15
	3	-1.42	-1.69 *	-0.89	-2.06 **	-1.27
	4	-3.67 ***	-3.84 ***	-3.19 ***	-4.16 ***	-3.45 ***
	5	-3.86 ***	-3.98 ***	-3.41 ***	-4.82 ***	-4.20 ***
	6	-4.02 ***	-4.03 ***	-3.90 ***	-4.93 ***	-4.70 ***
Year Lead	1	-1.30	-1.55	-0.40	-1.40	-0.15
	2	-3.67 ***	-3.84 ***	-3.19 ***	-4.16 ***	-3.45 ***
	3	-4.02 ***	-4.03 ***	-3.90 ***	-4.93 ***	-4.70 ***

Notes: The table reports results 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 $dDff_{t-1}$

Horizon		Model (d)	(e)	(f)	(i)	(j)
Quarter Lead	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 ***
Half Year Lead	1	0.94	0.96	0.92	1.41	1.50
	2	-0.94	-0.92	-0.99	-0.49	-0.42
	3	-0.62	-0.63	-0.71	-0.85	-0.72
	4	-2.78 ***	-2.70 ***	-2.78 ***	-2.75 ***	-2.55 **
	5	-2.83 ***	-2.81 ***	-2.83 ***	-3.32 ***	-3.02 ***
	6	-3.37 ***	-3.26 ***	-3.27 ***	-3.62 ***	-3.45 ***
Year Lead	1	-0.94	-0.92	-0.99	-0.49	-0.42
	2	-2.78 ***	-2.70 ***	-2.78 ***	-2.75 ***	-2.55 **
	3	-3.37 ***	-3.26 ***	-3.27 ***	-3.62 ***	-3.45 ***

Notes: The table reports results 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 dff_{t-1}

	Model																			
	(a)				(b)				(c)				(g)				(h)			
	TEST	Sig. level			TEST	Sig. level			TEST	Sig. level			TEST	Sig. level			TEST	Sig. level		
Variable		ASY	BSK	BS		ASY	BSK	BS		ASY	BSK	BS		ASY	BSK	BS		ASY	BSK	BS
dff_{t-1}	5.62	***			8.09	***		*	2.43	**			14.32	***	***	*	4.61	***		
$graym_t$	0.85				1.08				0.42				0.48				0.47			
$gray0_t$	1.88	*			0.25				0.63				1.08				1.98	*	*	
$gray1_t$	0.96				1.49				0.23				0.68				0.66			
$gray2_t$	1.49				1.63				0.38				0.35				0.22			
$igrym_t$	0.86				0.30				0.57				0.88				0.47			
$igry0_t$	0.71				0.30				0.47				7.09	***	***	**	7.05	***	***	***
$igry1_t$	1.27				0.48				0.61				1.61			*	1.23			**
$igry2_t$	2.55	**			0.57				0.59				3.08	**	**		3.87	***	***	**
$gradm_t$	0.74				1.83	*			1.78	*			4.78	***			2.96	**		
$grad0_t$	1.35				0.95				0.78				3.16	**			2.31	**		
$grad1_t$	1.42				2.56	**			2.70	**	*		3.67	***			3.26	***		
$grad2_t$	0.59				0.81				1.42				2.47	**			1.83	*		
$igrdm_t$	0.86				0.30				0.57				0.88				0.47			
$igrd0_t$	0.23				0.31				0.27				0.20				0.26			
$igrd1_t$	1.03				1.77	*			0.78				2.76	**			1.47			
$igrd2_t$	2.97	**	*		0.99				1.23				1.84	*	*		1.53			
$innovation_t$	1.19				2.44	**	**		0.20				2.01	*			0.97			
gdp_{t-1}	1.15				1.09				0.85				0.45				0.32			
gdp_{t-2}	0.67				1.17				0.86				0.27				1.09			
gdp_{t-3}	0.40				0.78				0.45				1.12				0.24			
gdp_{t-4}	0.49				0.48				1.98	*			0.48				0.39			
gdp_{t-5}	2.74	**	*		1.52				0.84				0.87				0.54			
gdp_{t-6}	1.15				0.65				1.50				0.61				1.40			
gdp_{t-7}	0.42				0.33				1.05				0.43				0.16			
gdp_{t-8}	0.41				0.34				0.17				0.48				0.66			
gdp_{t-9}	1.44				0.74				0.76				0.40				0.36			
Inf_{t-1}	1.24				0.64				0.59				0.45				0.31			
Inf_{t-2}	0.20				0.35				0.19				0.31				0.26			
Inf_{t-3}	0.18				0.59				1.03				0.69				0.95			
Inf_{t-4}	2.54	**			1.26				2.10	*	*		0.43				0.56			
Inf_{t-5}	0.80				0.67				0.72				1.11				0.57			
Inf_{t-6}	1.32				0.40				0.53				0.86				0.16			
Inf_{t-7}	0.24				0.63				0.22				0.37				0.10			
Inf_{t-8}	1.87	*	*		0.58				0.85				1.08				0.29			
Inf_{t-9}	0.51				0.41				0.35				0.45				0.11			

Notes: The table reports results for the semiparametric causality tests based on the moment condition (8) with $\phi(z_{t,i},v_2)$ equal to $1\{z_{t,i} \leq v_2\}$. Each line uses the specified variable as $z_{t,i}$. 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. Critical values use an asymptotic approximation (ASY), a bootstrap of the transformed test statustic (BSK), and a bootstrap of the untransformed statistic (BS). See text for details.

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

Table 2b: Specification Tests for models using $dDff_{t-1}$

	Model																			
	(d)				(e)				(f)				(i)				(j)			
	TEST	Sig. level			TEST	Sig. level			TEST	Sig. level			TEST	Sig. level			TEST	Sig. level		
Horizon		ASY	BSK	BS		ASY	BSK	BS		ASY	BSK	BS		ASY	BSK	BS		ASY	BSK	BS
$dDff_{t-1}$	8.13	***	**		14.49	***			10.39	***	***		2.22	**			5.34	***	**	
$graym_t$	1.18				0.36				1.70	*			0.18				0.20			
$gray0_t$	2.94	**	**		2.08	*			2.36	**			1.84	*			2.07	*		
$gray1_t$	1.82	*			1.10				0.54				0.56				0.58			
$gray2_t$	2.88	**			1.03				0.37				0.48				0.50			
$igrym_t$	2.03	*			1.23				0.65				0.98				0.97			
$igry0_t$	0.98				0.45				0.48				7.27	***	**	*	7.10	***	***	*
$igry1_t$	1.87	*			1.35				2.84	**			2.24	**	**	*	2.30	**	**	*
$igry2_t$	1.73	*			1.37				0.71				3.66	***	***	*	3.75	***	***	*
$gradm_t$	0.98				2.90	**			3.04	**			3.48	***			3.71	***		
$grad0_t$	1.11				0.87				1.59				2.39	**			2.72	**		
$grad1_t$	1.07				2.72	**			2.13	**			3.45	***			3.92	***		
$grad2_t$	0.52				1.36				1.42				2.16	**			2.64	**		
$igrdm_t$	2.03	*			1.23				0.65				0.98				0.97			
$igrd0_t$	0.58				0.70				1.08				0.30				0.28			
$igrd1_t$	0.79				0.54				0.36				1.90	*			1.83	*		
$igrd2_t$	5.74	***	**		2.09	*			3.03	**	*		1.95	*	*		1.90	*	*	
$innovation_t$	3.57	***			1.14				0.71				0.79				0.72			
gdp_{t-1}	1.41				0.85				1.30				0.34				0.28			
gdp_{t-2}	0.73				1.23				1.49				0.26				0.28			
gdp_{t-3}	1.14				0.86				1.87	*			0.46				0.55			
gdp_{t-4}	0.59				0.51				2.07	*			0.43				0.38			
gdp_{t-5}	3.69	***	*		1.86	*			1.37				0.64				0.49			
gdp_{t-6}	0.58				0.73				0.33				0.41				0.45			
gdp_{t-7}	0.40				1.12				0.57				0.22				0.20			
gdp_{t-8}	0.82				0.72				0.79				0.70				0.80			
gdp_{t-9}	2.67	**			0.98				1.57				0.44				0.40			
Inf_{t-1}	2.50	**			1.38				1.17				0.39				0.48			
Inf_{t-2}	0.43				0.23				0.39				0.51				0.54			
Inf_{t-3}	0.41				0.42				0.28				0.97				1.04			
Inf_{t-4}	2.89	**			2.18	**	*		1.65				0.33				0.40			
Inf_{t-5}	1.99	*			1.09				2.15	**			1.35				1.15			
Inf_{t-6}	1.73	*			0.72				1.03				0.87				0.70			
Inf_{t-7}	0.28				0.40				0.68				0.18				0.12			
Inf_{t-8}	3.64	***	*		0.96				1.27				0.74				0.61			
Inf_{t-9}	0.42				0.35				0.21				0.25				0.26			

Notes: The table reports results for the semiparametric causality tests based on the moment condition (8) with $\phi(z_{t,i},v_2)$ equal to $1\{z_{t,i} \leq v_2\}$. Each line uses the specified variable as $z_{t,i}$. 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. Critical values use an asymptotic approximation (ASY), a bootstrap of the transformed test statustic (BSK), and a bootstrap of the untransformed statistic (BS). See text for details.

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

Table 3a: Semiparametric Causality Tests using dff_{t-1}, for Up and Down Policy Changes

		Model																			
		(a)				(b)				(c)				(g)				(h)			
		TEST	Sig. level			TEST	Sig. level			TEST	Sig. level			TEST	Sig. level			TEST	Sig. level		
Horizon			ASY	BSK	BS		ASY	BSK	BS		ASY	BSK	BS		ASY	BSK	BS		ASY	BSK	BS
Quarters Lead	1	0.422				0.270				0.625				0.145				0.137			
	2	0.303				0.170				0.347				0.445				0.179			
	3	0.861				0.521				0.429				0.189				0.328			
	4	1.728	*			1.076				0.809				0.548				0.730			
	5	1.393				0.640				0.393				0.438				0.885			
	6	1.624				0.831				0.284				0.640				1.081			
	7	2.650	**	**		1.499				0.301				1.181				1.075		*	
	8	2.158	**	*	*	1.606			*	1.286				1.965	*	*	**	2.616	**	***	*
	9	1.618			*	1.423			*	1.783	*	*		2.771	**	**	**	2.564	**	**	*
	10	1.719	*	*		2.554	**	**	*	2.954	**	**		3.149	**	**	**	2.371	**	***	*
	11	3.058	**	**	*	4.619	***	***	**	3.049	**	**		3.725	***	**	**	3.220	***	***	**
	12	2.196	**	*	*	2.400	**	**	**	2.971	**	**	*	3.395	***	**	**	3.199	***	***	**
Half Year Lead	1	0.303				0.170				0.347				0.445				0.179			
	2	1.728	*			1.076				0.809				0.548				0.730			
	3	1.624				0.831				0.284				0.640				1.081			
	4	2.158	**	*	*	1.606			*	1.286				1.965	*	*	**	2.616	**	**	*
	5	1.719	*	*		2.554	**	*		2.954	**	**		3.149	**	**	**	2.371	**	***	
	6	2.196	**	*	*	2.400	**	*	**	2.971	**	**	*	3.395	***	**	**	3.199	***	***	**
Year Lead	1	1.728	*			1.076				0.809				0.548				0.730			
	2	2.158	**	*	*	1.606			*	1.286				1.965	*	*	**	2.616	**	**	*
	3	2.196	**	*	*	2.400	**	**	**	2.971	**	**	*	3.395	***	**	**	3.199	***	***	**

Notes: The table reports results for the semiparametric causality tests based on the moment condition (7) with $\phi(U_t,v)$ equal to $1\{y_t \leq v_1\}$. In this implementation, $\mathcal{D}_t$ 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. Critical values use an asymptotic approximation (ASY), a bootstrap of the transformed test statustic (BSK), and a bootstrap of the untransformed statistic (BS). See text for details.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3b: Semiparametric Causality Tests using $dDff_{t\text{-}1}$, for Up and Down Policy Changes

		Model																			
		(d)				(e)				(f)				(i)				(j)			
		TEST	Sig. level			TEST	Sig. level			TEST	Sig. level			TEST	Sig. level			TEST	Sig. level		
Horizon			ASY	BSK	BS		ASY	BSK	BS		ASY	BSK	BS		ASY	BSK	BS		ASY	BSK	BS
Quarters Lead	1	0.647				0.520				0.638				0.265				0.282			
	2	0.560				0.321				0.408				0.624				0.596			
	3	1.971	*			1.510				1.771	*			0.602				0.562			
	4	2.680	**			2.045	*			2.241	**	*		1.400		*		1.262		*	
	5	2.475	**			1.493				1.575				0.783				0.657			
	6	2.723	**	*		1.430				1.570				0.974				0.797			
	7	3.743	***	*		1.832	*			1.778	*			1.555		*		1.425		*	
	8	2.868	**	**		1.799	*	*		1.829	*	*	*	2.640	**	**		2.394	**	**	
	9	2.030	*			1.590		*		2.008	*	*		2.361	**	**		2.322	**	**	
	10	1.497				1.527				2.706	**	**		1.970	*	**		1.938	*	**	
	11	2.724	**	*		2.778	**	**		3.763	***	**		2.609	**	**	*	2.542	**	**	*
	12	2.461	**	*		2.387	**	**		3.475	***	**		2.610	**	***	*	2.537	**	***	*
Half Year Lead	1	0.560				0.321				0.408				0.624				0.596			
	2	2.680	**	*		2.045	*			2.241	**			1.400		*		1.262		*	
	3	2.723	**	*		1.430				1.570				0.974				0.797			
	4	2.868	**	*		1.799	*	*		1.829	*	*		2.640	**	**		2.394	**	**	
	5	1.497				1.527		*		2.706	**	**		1.970	*	**		1.938	*	**	
	6	2.461	**	*		2.387	**	**		3.475	***	**		2.610	**	***	*	2.537	**	**	
Year Lead	1	2.680	**			2.045	*			2.241	**			1.400		*		1.262			
	2	2.868	**	**		1.799	*	*		1.829	*	*		2.640	**	**		2.394	**	**	
	3	2.461	**	*		2.387	**	**		3.475	***	**		2.610	**	***	*	2.537	**	***	*

Notes: The table reports results for the semiparametric causality tests based on the moment condition (7) with $\phi(U_t,v)$ equal to $1\{y_t \leq v_1\}$. In this implementation, $\mathcal{D}_t$ 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. Critical values use an asymptotic approximation (ASY), a bootstrap of the transformed test statustic (BSK), and a bootstrap of the untransformed statistic (BS). See text for details.

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

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

Horizon		(a) TEST	(a) ASY	(a) BSK	(a) BS	(c) TEST	(c) ASY	(c) BSK	(c) BS	(e) TEST	(e) ASY	(e) BSK	(e) BS	(h) TEST	(h) ASY	(h) BSK	(h) BS
Quarters Lead	1	0.277				0.512				0.298				0.106			
	2	0.082				0.296				0.192				0.063			
	3	0.046				0.268				0.133				0.107			
	4	0.546				0.634				0.460				0.348			
	5	0.210				0.234				0.239				0.083			
	6	0.111				0.073				0.202				0.097			
	7	0.334				0.147				0.247				0.095			
	8	0.377				0.826				0.284				0.189			
	9	0.118				1.340	**	*		0.163				0.526			
	10	0.972	*	*		2.729	***	**		0.877				0.809		*	
	11	1.610	**	**		2.309	***	**		1.206	*	*		0.757		*	
	12	0.708				1.976	**	**		0.992	*			0.400			
Half Year Lead	1	0.082				0.296				0.192				0.063			
	2	0.546				0.634				0.460				0.348			
	3	0.111				0.073				0.202				0.097			
	4	0.377				0.826				0.284				0.189			
	5	0.972	*	*		2.729	***	**		0.877		*		0.809		**	
	6	0.708				1.976	**	**		0.992	*	*		0.400			
Year Lead	1	0.546				0.634				0.460				0.348			
	2	0.377				0.826				0.284				0.189			
	3	0.708				1.976	**	**		0.992	*	*		0.400			

Notes: The table reports results for the semiparametric causality tests based on the moment condition (7) with $\phi(U_t,v)$ equal to $1\{y_t \leq v_1\}$. In this implementation, $\mathcal{D}_t$ 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. Critical values use an asymptotic approximation (ASY), a bootstrap of the transformed test statustic (BSK), and a bootstrap of the untransformed statistic (BS). 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

		Model															
		(a)				(c)				(e)				(h)			
		TEST	Sig. level			TEST	Sig. level			TEST	Sig. level			TEST	Sig. level		
Horizon			ASY	BSK	BS		ASY	BSK	BS		ASY	BSK	BS		ASY	BSK	BS
Quarters Lead	1	0.145				0.113				0.222				0.031			
	2	0.220				0.051				0.129				0.116			
	3	0.815				0.161				1.377	**	*		0.221			
	4	1.182	*	*		0.176				1.585	**	*		0.382			
	5	1.183	*	*		0.159				1.255	*	*		0.803		*	
	6	1.513	**	**		0.210				1.228	*	*		0.984	*	*	
	7	2.316	***	**		0.154				1.585	**	*		0.980	*	**	
	8	1.780	**	**	**	0.459			**	1.516	**	**	*	2.427	***	***	**
	9	1.500	**	**	*	0.443			*	1.427	**	**	*	2.038	**	***	**
	10	0.747			*	0.225				0.650				1.563	**	**	*
	11	1.448	**	**	**	0.739		*	**	1.573	**	**	*	2.463	***	***	**
	12	1.488	**	**	**	0.995	*	**	**	1.394	**	**	*	2.798	***	***	**
Half Year Lead	1	0.220				0.051				0.129				0.116			
	2	1.182	*	*		0.176				1.585	**	*		0.382			
	3	1.513	**	**		0.210				1.228	*	*		0.984	*	*	
	4	1.780	**	**	**	0.459			*	1.516	**	**	*	2.427	***	***	**
	5	0.747			*	0.225				0.650				1.563	**	***	*
	6	1.488	**	**	**	0.995	*	**	**	1.394	**	**	*	2.798	***	***	**
Year Lead	1	1.182	*	*		0.176				1.585	**	*		0.382			
	2	1.780	**	**	**	0.459			**	1.516	**	**	*	2.427	***	***	**
	3	1.488	**	**	**	0.995	*	**	**	1.394	**	**	*	2.798	***	***	**

Notes: The table reports results for the semiparametric causality tests based on the moment condition (7) with $\phi(U_t, v)$ equal to $1\{y_t \leq v_1\}$. In this implementation, $\mathcal{D}_t$ 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. Critical values use an asymptotic approximation (ASY), a bootstrap of the transformed test statustic (BSK), and a bootstrap of the untransformed statistic (BS). See text for details.

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

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.

Andrews, D. W. (1995): "Nonparametric Kernel Estimation for Semiparametric Models," *Econometric Theory*, pp. 560–596.

Arcones, M. A., and B. Yu (1994): "Central Limit Theorems for Empirical and U-Processes of Stationary Mixing Sequences," *Journal of Theoretical Probability*, pp. 47–71.

Bai, J. (2002): "Testing Parametric Conditional Distributions of Dynamic Models," *mimeo*.

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 (2005): "Distribution Free Specification Tests of Conditional Models," *manuscript*.

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.

Hall, P., and C. Heyde (1980): *Martingale Limit Theory and its Application.* Academic Press.

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 (2003a): "Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score," *Econometrica*, 71, 1161–1189.

Hirano, K., G. Imbens, and G. Ridder (2003b): "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.

Pakes, A., and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57(5), 1027–1057.

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.

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.

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*.

Sute, 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.

VanderVaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes*. Springer Verlag.