NBER WORKING PAPER SERIES

SEMIPARAMETRIC CAUSALITY TESTS
USING THE POLICY PROPENSITY SCORE

Joshua D. Angrist
Guido M. Kuersteiner

Working Paper 10975
http://www.nber.org/papers/w10975

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2004

We thank Christina Romer for sharing the Romers' data, Alberto Abadie, Xiaohong Chen, Simon Gilchrist, Stefan Hoderlein 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 grant SES-0095132. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Semiparametric Causality Tests Using the Policy Propensity Score
Joshua D. Angrist and Guido M. Kuersteiner
NBER Working Paper No. 10975
December 2004
JEL No. C14, C22, E52

ABSTRACT

Time series data are widely used to explore causal relationships, typically in a regression framework with lagged dependent variables. Regression-based causality tests rely on an array of functional form and distributional assumptions for valid causal inference. This paper develops a semi-parametric test for causality in models linking a binary treatment or policy variable with unobserved potential outcomes. The procedure is semiparametric in the sense that we model the process determining treatment -- the policy propensity score -- but leave the model for outcomes unspecified. This general approach is motivated by the notion that we typically have better prior information about the policy determination process than about the macro-economy. 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. We also develop a test for full conditional independence, in contrast with the usual focus on mean independence. Our approach is illustrated using data from the Romer and Romer (1989) study of the relationship between the Federal reserve's monetary policy and output.

Joshua D. Angrist
Department of Economics
MIT, E52-353
50 Memorial Drive
Cambridge, MA 02142-1347
and NBER
angrist@mit.edu

Guido M. Kuersteiner
Department of Economics
Boston University
270 Bay State Road
Boston, MA 02215
gkuerste@bu.edu

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 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 true randomized trial or a compelling exclusion restriction. The analogy between a time series causal inquiry and a cross-sectional selection-on-observables argument is even stronger when the policy variable can be coded as a binary treatment. For example, we can consider the causal effect of exposure to a discrete monetary shock, with the latter viewed as a binary treatment. This is the essence of the approach taken in Romer and Romer's (1989) seminal analysis of the federal reserve's open market committee decisions, an application that we use here to illustrate theoretical ideas.

While providing a flexible tool for the analysis of causal relationships, an important drawback of regression-based conditional independence tests is that they typically require an array of auxiliary assumptions that are hard to assess and interpret, especially in a time series context. In addition to the linearity implicit in any regression test, researchers must choose conditioning variables, lag lengths, and impose assumptions that imply some sort of stationarity. The principal contribution of this paper is to develop an alternative approach to time series causality testing that shifts the focus away from modelling the relatively mysterious process determining outcomes towards a model of the process determining policy decisions. That is, we develop tests for causality that rely on a model for the conditional probability of treatment, which we call the "policy propensity score", leaving the model for outcomes unspecified. This approach seems especially appealing for the sort of time series applications we have in mind. In many of these cases there is some agreement – and even some evidence – as to what the conditioning variables used by policy makers are. Moreover, the binary nature of some policy variables provides a natural guide as to the choice of functional form. 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.

We use the time series causal framework to develop new distribution-free Kolmogorov-Smirnov (KS) and von Mises (VM) statistics that test for full conditional independence in time series models. The tests developed here are distribution-free in the sense that critical values do not depend on the sample for a given model design. Testing for full independence is also an innovation since most previous work on time series causality testing is concerned solely with mean independence. Finally, the tests developed here are semiparametric in the sense that a parametric model is used for the policy propensity score, but other features of the data-generating process are left unspecified. Our approach is related to earlier work on semiparametric estimation of average causal effects by Robins, Mark, and Newey (1992), who focus on sequential randomized trials. Also related is the Linton and Gozalo (1999) study of non-parametric causality tests in a cross-sectional context. Linton and Gozalo consider KS- and VM-type statistics, as we do, but the limiting distributions of their test statistics are not asymptotically distribution-free. These distributions are also difficult to bootstrap in a time series context. More recently, Su and White (2003) have proposed 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.

The main advantage of using the propensity score in our context lies in the fact that this reduces the problem of testing for conditional distributional independence to a problem of testing for a martingale difference sequence property of a certain function of the data. This problem is relatively easy to handle and 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). 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. 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.[1] Our work extends Koul and Stute (1999) by allowing for more general forms of dependence, including mixing and conditional heteroskedasticity. These extensions are important in our application because even under the null hypothesis of no causal relationship, the observed time series are not Markovian and do not have a martingale difference structure. Most importantly, direct application of the Khmaladze (1988,1993) method in a multivariate context appears to work poorly in practice. We therefore use a Rosenblatt (1952) transformation of the data in addition to the Khmaladze transformation. This combination of

[1] The univariate version of the Khmaladze transform was first used in econometrics by Bai (2002) and Koenker and Xiao (2002) .

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 and provides a heuristic derivation of our semiparametric test statistics. Strategies for constructing feasible versions of these statistics are discussed in Section 3 and Section 4 discusses the construction of feasible critical values. Although in principal straightforward, in practice, the distribution theory is complicated by the need to account for estimation of the propensity score.[2] We briefly explore finite-sample properties of the new statistics in a Monte Carlo study discussed in Section 5. The empirical behavior of alternative causality concepts and test statistics is illustrated through a re-analysis of the Romer and Romer (1989, 1994) data in Section 6. The last section of the paper concludes and suggests directions for further theoretical work.

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. See, e.g. Rosenbaum and Rubin (1983), who introduced the propensity score as a tool for causal inference in the potential-outcomes framework.

Our basic definition of causality relies on distinguishing the outcomes that would be realized with and without treatment, 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.

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 vector stochastic process $\chi_t(\omega) = (Y_t(\omega), X_t(\omega), D_t(\omega))$, defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $Y_t(\omega)$ is a vector of outcome variables, $D_t(\omega)$ is a vector of policy variables, and $X_t(\omega)$ is a vector of other exogenous and (lagged) endogenous variables that are not part of the null hypothesis of no causal effect of $D_t(\omega)$. The elements of the sample space, ω, can be thought of as indexing parallel universes, while the random variables defined on the sample space pick out the time series determined by realizations of ω. We assume that D_t takes values in the set $\mathcal{D}_t$. The observed sample χ_t is a realization of $\chi_t(\omega)$. Let $\bar{X}_t = (X_t, ..., X_{t-k}, ...)$ denote the covariate path, with similar definitions for $\bar{Y}_t$ and $\bar{D}_t$.

[2] Recent studies of the consequences of using an estimated propensity score for cross-sectional causal inference include Heckman, Ichimura and Todd (1998) Hahn (1999) and Hirano, Imbens, and Ridder (2003).

This framework leads to a definition of counterfactual outcomes based on the notion that past policy actions have the potential to change any future outcome variable, for each realization of the outcome ω :

Definition 1 *Assume that there exists a measurable map ξ_{t+j} such that*

$$Y_{t+j}(\omega) = \xi_{t+j}(\omega, D_t(\omega)) \text{ for all } t, j > 0 \text{ and almost all } \omega.$$

Potential outcomes are defined as

$$Y^d_{t+j}(\omega) = \xi_{t+j}(\omega, d) \text{ for all } d \in \mathcal{D}_t.$$

The sharp null hypothesis of no causal effects for potential outcomes is $Y^{d'}_{t+j}(\omega) = Y^d_{t+j}(\omega)$, $j > 0$ for all possible realizations $d, d' \in \mathcal{D}_t$. This coincides with the hypothesis of no causal effects in the simple situation studied by Rubin (1974), where we would write $Y_{0t} = Y_{1t}$.[3]

Our approach to causal testing allows the map ξ_{t+j} to be unspecified. On the other hand, it is common practice in econometrics to model χ_t as a function of its own lags and possibly exogenous variables or innovations in variables, and so it is worth thinking about what potential outcomes would be in this case. Given such a functional relationship, the map ξ_{t+j} can be constructed in an obvious way; a simple but common example is given below:

Example 1 *Suppose that $Y_t(\omega) = \sum_{k=0}^{\infty} \psi_k D_{t-k}(\omega)$, using notation that makes the dependence of the policy variables on the sample space explicit. This model could be one equation from a structural VAR. In this simple example, the map ξ_{t+j} is given by*

$$\xi_{t+j}(\omega, d) = \textstyle\sum_{k=0,k\neq j}^{\infty} \psi_k D_{t+j-k}(\omega) + \psi_j d.$$

Equivalently, $Y^d_{t+j}(\omega) = Y_{t+j}(\omega) + \psi_j(d - D_t(\omega))$. The sharp null hypothesis of no causal effect holds if and only if $\psi_j = 0$ for all j. This is the familiar restriction that the impulse response function be identically equal to zero.

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.

[3] 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.

As part of this setup, 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.

Condition 1 *Selection on observables:*

$$Y_{t+j}^{d}(\omega) \perp D_t(\omega) | \mathcal{F}_t \textit{ for all } j > 0 \textit{ and for all } d \in \mathcal{D}_t.$$

Note that implicit in this assumption is the notion that even after conditioning on observables, there is stochastic variation in policy decisions. This variation is taken to be due to idiosyncratic factors such as those detailed for monetary policy by Romer and Romer (2004). These factors include the variation over time in operating procedures used to convert information into decisions, changes in policy-makers' beliefs about the workings of the economy, decision-makers' tastes and goals, political factors, the temporary pursuit of objectives other than changes in the outcomes of interest (e.g., monetary policy that targets exchange rates instead of inflation or unemployment), and finally harder-to-quantify factors such as the mood and character of decision-makers. A key element of Condition 1 is that, conditional on observables, this idiosyncratic variation is taken to be independent of potential future outcomes.

Substituting using $Y_{t+j}^{d'}(\omega) = Y_{t+j}^{d}(\omega)$, the key testable conditional independence assumption can now be written in terms of observable distributions as:

$$Y_{t+1}, ..., Y_{t+j}, ... \perp D_t | \mathcal{F}_t. \tag{1}$$

In other words, conditional on observed covariates and lagged outcomes, there should be no relationship between treatment and outcomes Of course, Condition 1 is a strong restriction. But this condition is imposed in the rational expectations models outlined by Lucas (1972) and Sims (1980). In particular, when there are no informational asymmetries between the public and monetary authorities these models also imply that Equation 1 holds. The following example describes another assignment mechanism that satisfies this condition:

Example 2 *Suppose policies depend on observed variables $\mathcal{F}_t$ through the function $D(\mathcal{F}_t, t)$, as well as an unobserved (to the econometrician) variable, ε_t. Policies are determined by $D_t = f(D(\mathcal{F}_t, t), \varepsilon_t)$, where f is a general mapping. For example, Shapiro (1994) postulates $D_t = \mathbf{1}\{z_t'\theta + \varepsilon_t > 0\}$ where ε_t is iid Gaussian. In this case $D(\mathcal{F}_t, t) = z_t'\theta$, $f(a, b) = \mathbf{1}\{a + b > 0\}$ and $\mathcal{D}_t = \{0, 1\}$. If ε_t is independent of $Y_{t+j}^{d}(\omega)$, Condition 1 is satisfied. This means we can view ε_t as essentially randomly assigned, with no direct effect on outcomes.*

Tests based on condition (1) can be seen as testing a generalized version of Sims causality. 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.[4] 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 $P(Y_{t+1}, D_t | \mathcal{F}_t)$ 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. To prove the nonequivalence theorem, it is enough to give a counterexample. We do this for linear Gaussian processes since discrete variables can be defined as functions of underlying linear indices.

Example 3 *Assume that the vector $\chi_t = (y_t, x_t, D_t)$ takes values in $\mathbb{R}^3$ and that χ_t has a representation in terms of an overidentified structural VAR where $y_t = bx_{t-1} + cD_{t-1} + \varepsilon_{yt}$, $x_t = fD_t + \varepsilon_{xt}$ and $D_t = \varepsilon_{Dt}$ where $\varepsilon_t = (\varepsilon_{yt}, \varepsilon_{xt}, \varepsilon_{Dt})$ is such that $\varepsilon_t\tilde{}N(0, I_3)$ and I_3 is the 3×3 identity matrix. The impulse response function of y_t is $y_t = \varepsilon_{yt} + b\varepsilon_{xt-1} + (c + bf)\,\varepsilon_{Dt-1}$. Sims non-causality holds if $c + bf = 0$ which occurs if $c = 0$ and either $b = 0$ or $f = 0$ or if $c = -bf$. On the other hand, Granger non-causality requires that $c = 0$. We therefore can have Sims non-causality but Granger causality when $c \neq 0$ and $c = -bf$. On the other hand, we have Granger non-causality but Sims causality when $c = 0$ and both b and f are non-zero*

[4] 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.

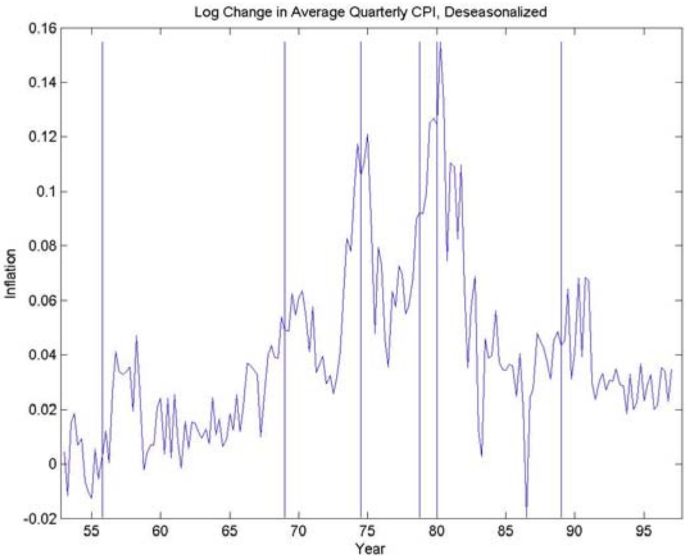

Figure 1: The vertical lines indicate Romer Dates.

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. This point is illustrated in Figure 1, which marks the Romer dates on the time series of inflation. In most cases, Romer dates are followed by an inflationary peak. This acceleration in inflation is both a cause of monetary policy and a response to earlier policy changes. Moreover, inflation may have effects on real variables. Thus, the causal relationship between monetary policy and activity in the real sector may be more appropriately analyzed in a framework that incorporates inflation and other nominal variables that respond to policy.

In the remainder of the paper, we assume the policy variable of interest is binary, although our conceptual framework applies more generally. We focus here on binary policy decisions because we are interesting in exploiting parallels with the cross-sectional treatment effects literature and because this leads naturally to a setup relying on the propensity score.[5] To develop this setup, we assume that models for the policy function can be written in the parametric form $P(D_t = 1|z_t) = p(z_t, \theta_0)$ for some function $p(.,.)$ and an unknown parameter vector, θ_0. Under the null hypothesis it follows that $P(D_t = 1|z_t, Y_{t+1}, ..., Y_{t+j}, ...) = P(D_t = 1|z_t)$. A 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. In the Monte Carlo and empirical parts of the paper, we explore simple Sims-type tests based on augmenting the policy function with future outcomes. But our main objective is to develop a more flexible class of semiparametric causality (conditional independence) tests that can be used to direct power in specific directions or to construct tests with power against general alternatives. A major advantage of our approach is that we do not have to attempt to identify and estimate a fully specified model of the entire macroeconomy or even the money-output relation. This saves the need to impose identifying restrictions on a complete structural VAR as in, e.g., Bernanke and Blinder (1992).

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 papers by the Romers and others inspired by their work.[6] The fact that D_t is binary in our application also suggests Logit or similar models provide a natural functional form. A motivating example that seems especially relevant in this context is Shapiro (1994), who develops a parsimonious Probit model of Fed decision-making as a function of net present value measures of inflation and unemployment. Finally, we note that while it is impossible to know for sure whether a given set of conditioning variables is adequate, diagnostic tests such as those proposed by Rosenbaum and Rubin (1985) can help decide when the model for the policy propensity score is an adequate representation of the role of the chosen set of covariates. A key technical advantage of reliance on the relatively tractable problem of modeling fed decision-making through the policy propensity score, is that this allows us to derive a semi-parametric test statistic with a

[5] 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).

[6] 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 (2004).

limiting distribution that depends only on the marginal distribution of outcome and conditioning variables (as opposed to the full joint distribution of the entire underlying process).

3 Semiparametric Causality Tests

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 binary, the conditional independence hypothesis can be written

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

We use the short hand notation $p(z_t) = P(D_t = i|z_t)$ and assume that $p(z_t) = p(z_t, \theta)$ is known up to a parameter θ.

Linton and Gozalo (1999) develop a fully nonparametric test of (4). Their test statistic is based on the empirical joint and marginal distributions of y_t, D_t, z_t. The resulting procedure is more flexible than ours but does not have a distribution-free limit distribution, a fact that leads Linton and Gozalo to bootstrap. In our setting, application of the bootstrap is complicated by the need to account for serial dependence and to impose the null while resampling. The bootstrap is also complicated by the fact that even under the null hypothesis the joint process of y_t, D_t, z_t is not Markovian and does not have a martingale difference sequence property. More recently, Su and White (2003) propose a nonparametric test based on estimates of conditional densities. Their procedure is asymptotically normal but converges more slowly than a $n^{-1/2}$ rate since their statistic involves non-parametric density estimates.

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

$$P(y_t \leq y, D_t = 1|z_t) - P(y_t \leq y|z_t)p(z_t) = E\left[\mathbf{1}(y_t \leq y)(D_t - p(z_t))|z_t\right] = 0. \tag{5}$$

This leads to a simple interpretation of test statistics based on this moment condition as looking for a relation between policy innovations, $D_t - p(z_t)$, and the distribution of future outcomes.

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 $\phi(.,.) : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$ be a function of U_t and some index v. Under the null we then have $E[\phi(U_t, v)(D_t - p(z_t))|z_t] = 0$. Examples 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). A natural choice for $\phi(U_t, v)$ is $\phi(U_t, v) = y_t^j \mathbf{1}\{z_t \leq v\}$ where $y_t^j = y_{1t}^{j_1} .... y_{k_1 t}^{j_k}$, which generates tests of conditional moment independence.

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, define the empirical process

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

with

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

This is similar in spirit to the process analyzed by Koul and Stute (1999) except for the fact that it depends on the parameter $v \in \mathbb{R}^k$ while Koul and Stute only consider the univariate case.[7]

Under regularity conditions that include stationarity of the observed process we show in Appendix A that $V_n(v)$ converges weakly to a limiting Gaussian process $V(v)$ on the space of cadlag functions[8] denoted by $\mathfrak{D}[-\infty, \infty]^k$ with covariance function $\Gamma(v, \tau)$, defined as

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

where $\nu, \tau \in \mathbb{R}^k$ and we note that $EV_n(v) = 0$. Using the fact that under the null $E[D_t|z_t, y_t] = E[D_t|z_t] = p(z_t)$ and partitioning $u = (u_1, u_2)$ with $u_2 \in [-\infty, \infty]^{k_2}$ we define $H(v \wedge \tau)$ with

$$H(v) = \int_{-\infty}^{v} \left(p(u_2) - p(u_2)^2\right) dF_u(u) \tag{6}$$

where $F_u(u)$ 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 $\Gamma(v, \tau) = \int \phi(u, v)\phi(u, \tau) dH(u)$. Note that when $\phi(U_t, v) = \mathbf{1}\{U_t \leq v\}$ then $\Gamma(v, \tau) = H(v \wedge \tau)$. This is the case we consider in the empirical application. The statistic $V_n(v)$ can be used to test the null hypothesis of conditional independence by comparing the value of $KS = \sup_v |V_n(v)|$ or $VM = \int (V_n(v))^2 dF_u(v)$ with the limiting distribution of these statistics under the null hypothesis.

Implementation of statistics based on $V_n(v)$ requires the construction of appropriate critical values. This problem is complicated by two factors affecting the limiting distribution of $V_n(v)$. The first factor is the dependence of $V_n(v)$ on $\phi(U_t, v)$ which induces data dependent correlation in the process $V_n(v)$. Hence, the nuisance parameter $\Gamma(v, \tau)$ appears in the limiting distribution. This is handled in two ways: First, critical values for the limiting distribution of $V_n(v)$ are computed numerically conditional on the sample in a way that accounts for the covariance structure $\Gamma(v, \tau)$. We discuss this procedure at the end of Section

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

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

4.1. An alternative to numerical critical values is to apply a transformation proposed by Rosenblatt (1952) which transforms $V_n(v)$ to a standard Gaussian process on the k-dimensional unit cube. The advantage of the latter transformation 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 and report critical values for the special case when $\phi(.,v) = \mathbf{1}\{. \leq v\}$ in Table 2.

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. Estimation of θ affects the limiting distribution of $\hat{V}_n(v)$ and needs to be taken into account. In Section 4 we discuss 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 $\hat{V}_n(v)$.

4 Implementation

Critical values for the KS and VM statistics are obtained through a series of transformations to correct for the fact that estimated parameters affect the relevant limiting distributions and to account for the correlation between the elements in U_t that lead to the presence of the nuisance parameter $\Gamma(v,\tau)$ in the limiting distribution.

As a first step, let $\hat{V}_n(v)$ denote the test statistic 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 7 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 $\bar{m}(v,\theta) = E\left[m(y_{t+k}, D_t, z_t, \theta; v)\right]$ 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

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

as shown in Appendix A. Next we turn to details of the transformations. Section 4.1 discusses a Khmaladze-type martingale transformation that corrects $\hat{V}(v)$ 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 $T\hat{V}(v)$ with the property that the transformed process, $W(v) = T\hat{V}(v)$, is a mean zero Gaussian process with covariance function $\Gamma(v, \tau)$. While $\hat{V}(v)$ has a complicated data-dependent limiting distribution (because of the estimated 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 $n^{-1/2}\sum_{t=1}^{n} l(D_t, z_t, \theta_0)$ 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 can not 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 $A_{-\infty} = \varnothing$, $A_\infty = [-\infty, \infty]^k$, $\lambda \leq \lambda' \implies A_\lambda \subset A_{\lambda'}$ and $A_{\lambda'} \backslash A_\lambda \to \varnothing$ as $\lambda' \downarrow \lambda$. Define the projection $\pi_\lambda f(v) = \mathbf{1}(v \in A_\lambda) f(v)$ and $\pi_\lambda^\perp = 1 - \pi_\lambda$ such that $\pi_\lambda^\perp f(v) = \mathbf{1}(v \notin A_\lambda) f(v)$. We then define the inner product $\langle f(.), g(.)\rangle := \int f(u)g(u)' dH(u)$ and the matrix

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

We note that the process $V(v)$ can be represented in terms of a Gaussian process $b(v)$ 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

$$T\hat{V}(v) := W(v) = \hat{V}(v) - \int \left\langle \phi(., v), d\left(\pi_\lambda \bar{l}(., \theta)\right)\right\rangle C_\lambda^{-1} \hat{V}(\pi_\lambda^\perp \bar{l}(., \theta)) \tag{7}$$

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

$$\bar{l}(v,\theta) = \frac{1}{(p(v_2,\theta) - p(v_2,\theta)^2)} \frac{\partial p(v_2,\theta)}{\partial \theta}.$$

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. [9]

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

$$\hat{W}_n(v) = T_n V_n\left(v\right) = \hat{V}_n\left(v\right) - \int \left(\int \phi(u,v) d\left(\pi_\lambda \bar{l}(u,\theta)\right) d\hat{H}_n(u) \right) \hat{C}_\lambda^{-1} \hat{V}_n(\pi_\lambda^\perp \bar{l}(.,\hat{\theta})) \tag{8}$$

with $\hat{V}_n(\pi_\lambda^\perp \bar{l}(.,\hat{\theta})) = n^{-1/2} \sum_{s=1}^n \pi_\lambda^\perp \bar{l}(U_s,\hat{\theta}) \left(D_s - p(z_s,\hat{\theta})\right)$ 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. Here we focus on sets $A_\lambda = [-\infty,\lambda] \times [-\infty,\infty]^{k-1}$, which turns out to be convenient in this context. Denote the first element of y_t by y_{1t}. Then (8) can be expressed more explicitly as

$$\hat{W}_n(v) = \hat{V}_n(v) - n^{-1/2} \sum_{t=1}^n \left[\phi\left\{U_t, v\right\} \frac{\partial p(z_t,\hat{\theta})}{\partial \theta'} \hat{C}_{y_{1t}}^{-1} n^{-1} \sum_{s=1}^n \mathbf{1}\left\{y_{1s} > y_{1t}\right\} \bar{l}(U_s,\hat{\theta}) \left(D_s - p(z_s,\hat{\theta})\right) \right] \tag{9}$$

Critical values for $\hat{W}_n(v)$ can be computed numerically as follows: Draw U_t^* randomly from the empirical distribution $\hat{F}_u(v)$. Let ε_t^* be an iid(0,1) random variable independent of U_t^*. Then

$$W_n^*(v) = n^{-1/2} \sum_{t=1}^n \varepsilon_t^* \mathbf{1}\left\{U_t^* \leq v\right\} \tag{10}$$

has the same limiting distribution as $\hat{W}_n(v)$ by standard arguments (see Van der Waart and Wellner, 1996). Critical values for $\hat{W}_n\left(v\right)$ can therefore be computed by repeatedly drawing from the distribution of $W_n^*(v)$. In Section 5 we report Monte Carlo results based on critical values obtained numerically from $W_n^*(v)$. These results show some size distortions. We therefore turn in the next section to a further transformation that leads to a distribution free limit for the test statistics.

[9] Stute, Thies and Zhu (1998) analyze a test of conditional mean specification in an independent sample allowing for heteroskedasticity by rescaling the equivalent of our $m(y_t, D_t, z_t, \theta_0; v)$ by the conditional variance. 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.

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 such as Table 2. Let $U_t = [U_{t1}, ..., U_{tk}]$ and define the transformation $w = T_R(v)$ component wise by $w_1 = F_1(v_1) = P(U_{t1} \leq v_1)$, $w_2 = F_2(v_2|v_1) = P(U_{t2} \leq v_2|U_{1t} = v_1)$,..., $w_k = F_k(v_k|v_{k-1}, ..., v_1) = P(U_{tk} \leq v_k|U_{tk-1} = v_{k-1}, ..., U_{t1} = v_1)$. 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 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

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

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 $E[D_t\phi(w_t, w)|z_t] = E[\phi(w_t, w)|z_t]\, p(z_t, \theta)$, or equivalently,

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

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

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

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. As a strategy for dealing with this arbitrariness, Justel, Pẽna and Zamar (1997) propose the use of tests d_i indexed by all possible $k!$ permutations of the elements of w_t and consider summary statistics such as $\max_i d_i$. We investigate the performance of this strategy in the Monte Carlo and empirical sections below.

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[10]

$$T_w\hat{V}_w(w) := W_w(w) = \hat{V}_w(w) - \int \left\langle \phi(., w), d\pi_\lambda \bar{l}_w(., \theta)\right\rangle C_\lambda^{-1} \hat{V}_w(\pi_\lambda^\perp \bar{l}_w(., \theta)). \qquad (11)$$

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

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

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

is a Gaussian process with covariance function $\int_0^1 \cdots \int_0^1 \phi(u,w)\phi(u,w')du$, where $h_w(.) = p(\left[T_R^{-1}(w_t)\right]_z, \theta)(1-p(\left[T_R^{-1}(w_t)\right]_z, \theta))$.

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 (9) 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

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

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

$$\hat{h}_w(.) = p(.,\hat{\theta})\left(1-p(.,\hat{\theta})\right).$$

The empirical version of the transformed statistic is

$$\begin{aligned}\hat{B}_{\hat{w},n}(w) &= \hat{W}_{\hat{w},n}\left(\phi(.,w)/\hat{h}_w(.)^{1/2}\right) \\ &= n^{-1/2}\sum_{t=1}^{n}\hat{h}_w(z_t)^{-1/2}\left[D_t - p(z_t,\hat{\theta}) - \hat{A}_{n,t}\right]\phi(\hat{w}_t,w) \qquad (12)\end{aligned}$$

where $\hat{A}_{n,s} = n^{-1}\sum_{t=1}^{n} \mathbf{1}\{\hat{w}_{t1} > \hat{w}_{s1}\}\left(D_t - p(z_t,\hat{\theta})\right)\frac{\partial p(z_s,\hat{\theta})}{\partial\theta'}\hat{C}^{-1}_{\hat{w}_{1s}}\bar{l}(z_t,\hat{\theta})$. 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 $\Upsilon_{[0,1]} = \left\{w \in [0,1]^k \,|w = \pi_x w\right\}$. 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

$$VM_w = \int_{\Upsilon_{[0,1]}} \left(\hat{B}_{\hat{w},n}(w)\right)^2 dw \Rightarrow \int_{\Upsilon_{[0,1]}} \left(B_w(w)\right)^2 dw \tag{13}$$

and

$$KS_w = \sup_{v\in\Upsilon_{[0,1]}} \left|\hat{B}_{\hat{w},n}(w)\right| \Rightarrow \sup_{v\in\Upsilon_{[0,1]}} \left|B_w(w)\right|. \tag{14}$$

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 (13) and the supremum in (14) can be computed over a discrete grid. The asymptotic representations (13) and (14) make it possible to use asymptotic statistical tables. For the purposes of the Monte Carlo below, we computed critical values for the VM statistic in the special case where $\phi(.,v) = \mathbf{1}\{. \leq v\}$ (These are reported in Table 2). These critical values depend only on the dimension k and are thus distribution free.[11] Table 2 is also used to construct critical values in our empirical application in Section 6.

5 Monte Carlo Evidence

We evaluated the performance of our semiparametric tests using a simple data generating process that nevertheless captures important features of the empirical applications we have in mind. The process is

$$\begin{aligned} y_t &= \beta y_{t-1} + \gamma D_t + \varepsilon_t \\ D_t &= \mathbf{1}\{y_{t-1} - \alpha + \eta_t > 0\}, \end{aligned}$$

where ε_t and η_t are independent with $\varepsilon_t \sim N(0,1)$ and η_t has a logistic distribution. We choose $\alpha = 3$ which leads to an unconditional probability of $D_t = 1$ of roughly 5% which is comparable to our empirical sample. This model has a standard lagged-dependent-variable structure that captures serial correlation in the outcome. The policy assignment is also correlated with lagged outcomes.

The simulation used samples of 100 and 200 observations in 500 replications. The reported results are rejection rates for the test statistics derived above and for a conventional t-test for the significance of γ in

[11] See Section 5 for a more detailed discussion of how Table 2 was constructed.

a regression of y_t on y_{t-1} and D_t. To construct the semiparametric test statistics, we used a Logit model for the propensity score and the test function $\phi(U_t, v) = \mathbf{1}\{(y_t, y_{t-1}) \leq v\}$.

Table 1 reports results for several implementations of our semiparametric test. These results are for the statistic $VM = \int \left(\hat{V}_n(v)\right)^2 dF_u(v)$ and differ only in the way in which $\hat{V}_n(.)$ is implemented and the method by which critical values were obtained. We choose a bandwidth of $m = 10n^{-1/(2+k)}$.[12]

We begin with statistics and significance levels calculated using the numerical methods to determine critical values, as described in Section 4.1. In particular, Column (1) in Table 1 reports results for the statistic $\hat{W}_n(v)$ defined by (9), with critical values obtained by numerical simulation conditional on the sample as described by equation (10). The test statistic reported here can therefore be written $\int \left(\hat{W}_n(v)\right)^2 d\hat{F}_u(v)$, where $\hat{F}_u(v)$ is the empirical distribution of U_t. This statistic relies on the Khmaladze transformation alone to adjust inference for estimation of the propensity score.

Test's based on the asymptotic critical values reported in Table 2 and using the Rosenblatt transformation as in (12), were constructed as follows. Let $d_i = \int \left[\hat{B}_{\hat{w},n}(w)\right]^2 dw$ be the statistic based on the i-th permutation of the elements in U_t before the Rosenblatt transform is applied to U_t. Column (2) in Table 1 reports results for the statistic $md \equiv \max_i d_i$ where the maximum statistic is taken over all permutations of the elements in U_t and uses critical values from Column (1) of Table 2.[13] We use the notation md_a to denote results for the md statistic that are based on asymptotic critical values.

Column (3) of Table 1 was calculated using upper bounds for the asymptotic critical values of the md statistic proposed by Justel, Peña and Zamar (1997). We use the notation md_b to denote results for the md statistic that are based on upper bounds. Upper bounds are based on $P(md > c_\alpha) \leq \sum_i P(d_i > c_\alpha) = k!\alpha$ such that $\alpha_{md} = \alpha/k!$ leads to a critical value with $P(md > c_{\alpha/k!}) \leq \alpha$. When $\alpha = .05$ is the desired significance level, we use the critical value corresponding to d for $k = 2$ and $1 - \alpha = .975$ in Table 2. Columns (4) and (5) of Table 1 report corresponding results based on asymptotic critical values for d_1 and d_2. Since in this case $U_t = \{y_t, y_{t-1}\}$ these two tests are based on the two permutations of U_t, $\{y_t, y_{t-1}\}$ and $\{y_{t-1}, y_t\}$.

As predicted by the theoretical discussion, the results in Table 1 show the tests d_1 and d_2 to have similar properties, with accurate size at all degrees of serial correlation in y_t that we investigated. When

[12] Experimentation with different choices of m indicate that the tests are not very sensitive to this parameter. Nevertheless, for much smaller values of m such as $m = n^{-1/(2+k)}/10$ we found that the tests were undersized.

[13] Note that the asymptotic critical values for d_i do not depend on the permutation chosen. For this reason we only distinguish between the maximum statistic md and d in Table 2. Critical values do depend on the dimension k of the vector U_t. Table 2 was obtained by randomly drawing U_t^{**} from a Gaussian distribution with a randomly drawn covariance matrix and then applying the Rosenblatt transform to the generated random variables U_t^{**}.Note that here the Rosenblatt transform T_R is known because U_t^{**} is Gaussian. We thus compute $d_i^{**} = n^{-1}\sum_{t=1}^{n} \varepsilon_t^* T_R^{-1}\left(U_t^{**}\right)$ for the i-th permutation of U_t^{**} where ε_t is iid standard Gaussian. The sample size is set to $n = 100$ and $100,000$ replications of d_i^{**} are used to approximate the distributions of md and d.

compared with a t-test, reported in Column (6), which in this scenario is both asymptotically optimal and has good finite sample size properties, the tests d_i fare quite well. It is especially encouraging to see that the semiparametric test statistics have good power properties, though these naturally fall somewhat short of the power for the parametric t-test.

The semiparametric tests have most accurate size when the asymptotic critical values for the statistic md are used, reported in Column (2). The resulting test is only slightly oversized for most values of β. Power is also quite good in this case, and the md_a test is at least as powerful as the individual statistics, d_i, although the differences are very small. This may be due in part to small size distortions of the md_a test. The md_b test based on upper bound critical values, reported in Column (3), is somewhat undersized for models with larger values of β and consequently has less power. This version of the test therefore leads to a conservative test of the null hypothesis.

Finally, the version of the test based on simulated critical values conditional on the sample in Column (1) has size distortions somewhat larger than the distortion for the individual tests d_i based on asymptotic critical values in Columns (4) and (5) when β is low to moderate, i.e. $\beta \leq .5$. At the same time, with simulated critical values, power is somewhat lower, a feature which clearly makes this implementation less attractive. Moreover, when $\beta = .9$ this version of the test displays fairly large size distortions, unlike the other implementations of the test. Overall, the md_a test using asymptotic critical values seems to provide the best combination of accuracy and power. The md_b test using upper bound critical values leads to a more conservative version of the test. We therefore used both test statistics for the empirical work described in the next section. At least in our application we found the differences to be minor with md_a only leading to slightly more significant results.

6 Causal Effects of Monetary Policy Shocks Revisited

In an influential study of the effects of monetary policy, Romer and Romer (1989) constructed a monetary policy shock variable derived using what they call the narrative approach, inspired by Friedman and Schwartz' classic monetary history. The narrative approach uses Federal Open Market Committee minutes to construct a dummy variable, D_t, to indicate episodes where the Fed took a marked anti-inflationary stance. Thus, D_t indicates periods that are now known as "Romer dates." The Romer dates mark Fed decisions to change short term interest rates, discount rates, or reserve requirements. There were six such dates in the original Romer sample, running from 1948-1987, with a 7th date added when the sample was extended through 1991 in Romer and Romer (1994). The link between Romer Dates and later economic activity provides a natural setting for propensity-score based estimates of the effects of monetary policy.

The key identifying assumption in the Romer papers is that, conditional on lagged outcomes, the Romer dates are as good as randomly assigned in the sense that regressions of future output growth on (lagged) dummies for these dates have a causal interpretation. A substantial literature has developed challenging this premise. Examples include Leeper (1997), who argues the Romer dates are determined in part by the Fed's (nonlinear) forecast of future output and Shapiro (1994) who similarly argues that monetary policy is forward-looking in a way that induces omitted variables bias in the Romers' regressions. Both of these critiques are consistent with the modeling strategy outlined here in that we focus attention on models for the policy-determination process. Romer and Romer (1997) defend the notion that, after appropriate conditioning, the dates can be seen as exogenous. Romer and Romer (2004) provide new estimates of the dates of monetary shocks using a somewhat more systematic version of the narrative approach. We focus on the original Romer dates because they correspond to our binary-policy-variable setup, though in future work we hope to address the more general policy evaluation problem.

Our re-analysis of the Romer data begins with Granger-style regressions of the growth of industrial production (IP) on contemporaneous and lagged dummies for Romer dates ("Romer dummies"), controlling for lags of IP growth. This is similar to the Romer's econometric approach, with two modifications. First, we aggregate monthly data to the quarterly level since there is probably little additional information in the higher-frequency series. Romer quarters are identified as quarters with a Romer month.[14] This also serves to increase the proportion of the sample coded as a Romer date, making it easier to estimate the policy propensity score. Second, the Romers assess the role of monetary policy variables by looking at the impulse response function, while we focus on F-statistics for the Romer dummies.

Controlling for 8 lags of output and no other covariates, a test for the joint significance of the Romer dummies generates a p-value of about .01. This result, consistent with the Romers' original findings, can be seen in the first two columns of Table 3.[15] We report both robust F-statistics based on White standard errors as well as non-robust standard errors. Significance levels using robust standard errors tend to be higher, especially in models with additional covariates. Non-robust standard errors may be more reliable in these cases since increased precision with robust standard errors is often an artifact of finite sample bias and size distortion (Chesher and Jewitt, 1987).

Much of the debate over the Romer's empirical approach focuses on whether it is enough to control for lagged output when assessing the causal effect of Romer dates on output. An especially important

[14] Quarterly series for all variables were constructed by averaging monthly series. Growth rates were constructed as the first differences of the log of the quarterly averages. All quarterly series were deseasonalized by recursive regressions on quarter dummies. The regressions are recursive in that coefficients were estimated using only information available prior to each observation. This procedure allows us to ignore the estimation error arising from this de-seasonalization. The series used in this section are listed in the last table. The original monthly series were obtained from the Wharton/DRI Global Insight service. Although standard and widely available, these series differ somewhat from the Romers' original as they have since been revised. We us the 1952-91 sample used by Shapiro (1994).

[15] The specification includes 12 lagged Romer dummies (3 years worth). This corresponds to the Romers' original equation which included 3 years worth of lagged Romer dummies.

control variable in this context is inflation, since the Fed presumably looks at this when making monetary policy decisions. On the other hand, inflation clearly responds to monetary policy and may therefore not be an exogenous control. This possibility was highlighted in the discussion of Granger-testing pitfalls in Example 3. To explore the consequences of adding inflation controls, we fit a version of the Romer's principal estimating equation after adding eight lags of inflation to the list of covariates. These results, reported in Columns 3 and 4 of Table 3, show that the addition of inflation controls reduces the significance level of the Romer dummies somewhat, though some effects are still significant. Similarly, adding controls for lagged unemployment rates further reduces the significance of the joint F test for the Romer dummies. These results appear in columns 5 and 6 of the table. The p-value for the joint significance of the Romer dummies becomes .09 for the non-robust version of the F-statistic.

Finally, we explore Sims-type semiparametric tests of conditional independence in this context using the transformed VM and KS statistics described above. For purposes of comparison, results from a parametric analog of the semiparametric tests are also reported. The semiparametric test results are for tests of conditional distributional independence where $\phi(U_t, v) = \mathbf{1}\{U_t \leq v\}$ and the policy propensity score was estimated using Logit, as for the Monte Carlos in the previous section. The semiparametric tests were implemented using the same bandwidth as used for the simulations.[16]

The foundation of our semiparametric testing procedure is a parsimonious model for the policy propensity score. Following Shapiro (1994), we used a parsimonious model based on the notion, also discussed by Romer and Romer (2004), that the systematic component of Fed policy decisions is driven by forecasts of inflation and unemployment. In particular, we first fit a vector autoregressive model (VAR) to unemployment and inflation. We then used predictions up to 100 periods ahead to construct a forecast of the "present value" of future inflation and unemployment in each period, similar to the present value forecasts used by Shapiro. The idea is that the Fed sets monetary policy based on this measure, or other summary forecasts that are highly correlated with this one. A detail here, however, is that because Shapiro's forecasting parameters were estimated on the entire sample, the resulting present value measures are not part of the relevant information set of the Fed. To avoid this conceptual (if not practical) problem, we also used a true out-of-sample forecasting procedure to construct the present value measures by estimating the VAR parameters on the sample prior to the forecast period only. The present value inflation and unemployment forecasts are the main covariates in the model for the policy propensity score, though some estimates also include lagged dependent variables.[17]

[16] We have experimented with different choices of the bandwidth, but found that the results are not sensitive to the choice of m.

[17] Lagged Romer dummies were also used as explanatory variables in the forecasting equations. The discount rate was set at 2%. The forecasting equation has eight lags for inflation and unemployment and 16 lags for the Romer dummies. When constructing out-of-sample forecasts, lag length for all covariates was reduced to four periods at the beginning of the sample.

The semiparametric md_b tests were constructed for three specifications, with results reported in columns 1-3 of Table 4. The first two specifications use full-sample and out-of-sample forecasts. The third specification adds lagged dependent variables to the model using out-of-sample forecasts. Results in different rows are for different lead lengths, e.g., causal effects on output growth one period ahead, two periods ahead, and so on. We look at each lead one at a time because the number of permutations required for the Rosenblatt transform grows rapidly with the dimensionality of a joint test. The upper bound method was used to obtain critical values for the semiparametric tests.[18]

Results from the first specification offer some evidence of a money-output relation at some lags. In particular, semiparametric tests reject non-causality at one, three to five, seven and eight quarter leads. These results may be misleading because full sample estimation of z_t invalidates the semiparametric tests. Significance levels are reduced considerably when out-of-sample forecasts are used to construct control variables, but there are still rejections at the third and eight quarter leads and a weakly significant result at the first lead. Adding lagged output growth does not change these findings, which can be seen in column 3. In Table 5 we report the same statistic but now judged against asymptotic critical values from Column (3) of Table 2. The results are essentially the same except that the first lead is now statistically significant. On balance, it seems fair to say that a forward-looking Sims test provides weak support the Romers' original conclusion, at least as far the correlation between Romer dates and future output growth as concerned. Not surprisingly, however, given the paucity of Romer dates that form the essence of the "natural experiment" that lies behind this inquiry, the evidence for money-output causality can fairly be described as "mixed."[19]

To gauge the extent to which our semiparametric test results have reduced power relative to similar parametric tests, we added future output growth to the present value variables (and possibly lagged output growth variables) already in the policy propensity score. The significance level of future output variables in the policy propensity score provides a parametric Sims-type test of a particular version of the conditional independence hypothesis that is at the heart of the semiparametric tests. In particular, we report significance levels for the coefficient θ_3 in the following choice equation of a Logit model for Fed

[18] The p-values reported in the table were obtained by translating the p-values in Table 2 into p-values for the upper bound test by multiplying α by $k!$. To achieve a 5% level of significance and with $k = 3$, this implies a critical value corresponding to $1 - \alpha = .9916$ needs to be used in Column (4) of Table 2. In Table 4 we report intervals for p-values. These are constructed translating the interval of critical values in which the test statistic falls into corresponding significance values. For example if $md = .4$ then the interval of critical values is $[.33, .42]$ from Column (4) of Table 2 with corresponding $\alpha \in [.01, .025]$. Because $k! = 6$ this translates into an effective α that is contained in $[.06, .15]$.

[19] The Romer's original findings showed statistical significance for Romer dummies at particular groups of lags in Granger-style regressions. These effects were large enough to induce a clear shift in the impulse response function, a relation analogous to the one checked by our forward-looking Sims tests.

action

$$D_t = 1\left\{\theta_0 + \theta_1 u_t^{pv} + \theta_2 \pi_t^{pv} + \theta_3 Y_{t+j} + \varepsilon_t > 0\right\}.$$

where $z_t = [u_t^{pv}, \pi_t^{pv}]'$ contains the present value measures for unemployment and inflation discussed earlier. The variable Y_{t+j} is the change in industrial production at lead j. Under the null hypothesis, the parameter θ_3 should be zero. The parametric version of this test has the advantage that, subject to having correctly specific the policy propensity score, the model is correct under the null hypothesis of non-causality. On the other hand, this specification need not be correct under the alternative, even if the policy propensity score is correctly specified, and may therefore have reduced power in some directions.

As it turns out, results from the parametric analog of our semiparametric tests are generally in line with the semiparametric results, especially for the out-of-sample forecast case. This can be seen in columns 4-6 of Table (4). In particular, there is some evidence of causality at the first and eighth lead, while two of the specifications also show something at an intermediate lead, in this case the third. The fact that the semiparametric and parametric models generate results with the same patterns of significance suggests power considerations do not substantially handicap the semiparametric results.

7 Conclusions and Directions for Further Work

This paper develops a causal framework for time series data using the notion of potential outcomes commonly used in cross-sectional evaluation research. This leads to a definition of causality similar to the one proposed by Sims. For models with covariates, Sims causality is not the same as Granger causality, and the distinction between these two concepts turns out to be conceptually important. In particular, Granger causality may confuse system dynamics with the causal effects of isolated policy actions. In contrast, Sims causality hones in on isolated policy shocks relative to a well-defined counter-factual baseline.

A major part of our agenda is to develop a causality test that focuses on the policy assignment mechanism, which we call the policy propensity score. In particular, we develop a new semiparametric test of conditional independence, valid under the selection-on-observables null hypothesis that is at the heart of much of the empirical work on time series causal effects. A major advantage of this approach is that it does not require researchers to model the process determining the outcomes of interest. The resulting tests have power against all alternatives but are necessarily joint tests of the null of no causality and correct specification of the policy propensity score.

The development here is limited to binary treatments but it seems likely our approach can be extended to multivalued treatments, perhaps along the lines explored in recent work by Hirano and Imbens (2004). Of course, it is an open question whether the technical machinery used here, such as the Khmaladze transform, transfers to the more general setting. This is a question we hope to address in future work.

We also plan to explore the question of whether tests for conditional mean and second-order moment independence have advantages over omnibus tests.

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 $\mathcal{A}_1^l = \sigma\left(\chi_1, ..., \chi_l\right)$ be the sigma field generated by $\chi_1, ..., \chi_l$. The sequence χ_t is β-mixing or absolutely regular if

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

A sequence is called α-mixing if

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

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

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

Condition 3 *Let $F_u(u)$ be the marginal distribution of U_t. Assume that $F_u\left(.\right)$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^k$ and has a density $f_u(u)$*

Condition 4 *The function $\phi(.,.)$ belongs to a VC subgraph class of functions with envelope $M(\chi_t)$ such that $E\left|M(\chi_t)\right|^{2+\delta} < \infty$ 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 5 *Let $H(v)$ be as defined in (6). 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 $h(v) = \partial^k H(v)/\partial v_1 ... \partial v_k$ and assume that $h(v) > 0$ for all $v \in \mathbb{R}^k$.*

Remark 1 *A sufficient condition for Condition (5) is that $0 < p(z_t, \theta_0) < 1$ almost surely.*

A.2 Limiting Distributions

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

Proposition 2 *Assume that Conditions 2, 3 and 5 are satisfied. Let $v_i \in \left[-\infty,\infty\right]^k$ for $i = 1,...,s$ be a finite collection of points. Then, for all finite s, $V_n\left(v_1\right),....,V_n\left(v_s\right)$ converges in distribution to a Gaussian limit with mean zero and covariance function $\Gamma(v_i,v_j)$. Moreover, $V_n\left(v\right)$ converges in $\mathfrak{D}\left[-\infty,\infty\right]^k$ to a Gaussian process $V(v)$ with covariance kernel $\Gamma(v,\tau)$ with $v,\tau \in \left[-\infty,\infty\right]^k$ and $V(-\infty) = 0$, $H(v)$ is positive 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 $E\left(m_v(\chi_t)|z_t\right) = 0$. Let $\lambda = (\lambda_1,...,\lambda_s)'$ with $\|\lambda\| = 1$. 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 $\varepsilon > 0$,

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

because $EY_{nt}^{2+\delta}$ is bounded for some $\delta > 0$. Also,

$$\begin{aligned}
\sum_t E\left[Y_{nt}^2|\mathcal{A}_1^{t-1}\right] &= n^{-1}\sum_{t=1}^n E\left[Y_t^2|\mathcal{A}_1^{t-1}\right] \\
&= n^{-1}\sum_{t=1}^n E\left[p(z_t,\theta_0)\left(1-p\left(z_t,\theta_0\right)\right)\left(\lambda_1\boldsymbol{\phi}\left(u_t,v_1\right)+\lambda_2\boldsymbol{\phi}\left(u_t,v_2\right)+...+\lambda_s\boldsymbol{\phi}(u_t,v_s)\right)^2|\mathcal{A}_1^{t-1}\right] \\
&\xrightarrow{p} \sum_{i,j}\lambda_i\lambda_j\Gamma(v_i,v_j)
\end{aligned}$$

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 $\hat{V}_n\left(v\right)$ evaluated at the estimated parameter value $\hat{\theta}$. The fact that $l\left(D_t,z_t,\theta_0\right)$ 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 6 *Let $\theta_0 \in \Theta$ where $\Theta \subset \mathbb{R}^d$ is a compact set and $d < \infty$. Assume that $E\left[D_t|z_t\right] = p(z_t|\theta_0)$ and for all $\theta \neq \theta_0$ it follows $E\left[D_t|z_t\right] \neq p(z_t|\theta)$. 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 $N_\delta\left(\theta_0\right)$ 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 $\partial p_i(z_t|\theta)/\partial\theta$ be the i-th element of the vector of partial derivatives $\partial p(z_t|\theta)/\partial\theta$ and let $\bar{l}_i(z_t,\theta)$ be the i-th element of $\bar{l}\left(z_t,\theta\right)$. Assume that there exists a function $B(x)$ and a constant $\alpha > 0$ such that*

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

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

$$E\left[\left(p(z_t,\theta_0)\left(1 - p(z_t,\theta_0)\right)\right)^{-(4+\delta)}\right] < \infty$$

and

$$E\left|\left(\partial p_i(z_t|\theta_0)/\partial\theta\right)^2 / \left(p(z_t,\theta_0)\left(1 - p(z_t,\theta_0)\right)\right)\right|^{\frac{4+\delta}{2}} < \infty$$

for all i and some $\delta > 0$.

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

Condition 7 *Let*

$$l\left(D_t, z_t, \theta\right) = \Sigma_\theta^{-1}\frac{\left(D_t - p(z_t,\theta)\right)\partial p\left(D_t|z_t,\theta\right)/\partial\theta}{p\left(D_t|z_t,\theta\right)\left(1 - p\left(D_t|z_t,\theta\right)\right)} \tag{15}$$

where

$$\Sigma_\theta = E\left[\frac{\partial\log p\left(D_t|z_t,\theta\right)/\partial\theta\partial\log p\left(D_t|z_t,\theta\right)/\partial\theta'}{p\left(D_t|z_t,\theta\right)\left(1 - p\left(D_t|z_t,\theta\right)\right)}\right]. \tag{16}$$

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

Proposition 3 *Assume that Conditions 2, 3,4, 5, 6 and 7 are satisfied. Then*

$$\sup_{v \in [-\infty,\infty]^k}\left\|\hat{V}_n\left(v\right) - V_n\left(v\right) + \dot{m}(v,\theta_0)n^{-1/2}\sum_{t=1}^n l\left(D_t, z_t, \theta_0\right)\right\| = o_p(1) \tag{17}$$

and if $l\left(D_t, z_t, \theta_0\right)$ is as defined in 15 and 16 then $\hat{V}_n\left(v\right)$ converges weakly in $\mathfrak{D}[-\infty,\infty]^k$ equipped with the sup norm to a limiting Gaussian process with mean zero and covariance function $\hat{\Gamma}(v,\tau) = \Gamma\left(v,\tau\right) - \dot{m}(v,\theta_0)L(\theta_0)\dot{m}(\tau,\theta_0)'$ where $L(\theta_0) = \Sigma_{\theta_0}^{-1}$ defined in 16.

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

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

where $\|\theta_n - \theta_0\| \leq \left\|\hat{\theta}-\theta_0\right\|$ by the mean value theorem. Let $\dot{m}(\theta,v) = E\left[\frac{\partial p(z_t,\theta)}{\partial\theta}\phi(U_t,v)\right]$ and $\dot{m}(U_t,\theta,v) = \frac{\partial p(z_t,\theta)}{\partial\theta}\phi(U_t,v) - \dot{m}(\theta,v)$. From Pakes and Pollard (1989, Lemmas 2.13 and 2.14) and Condition (6) it follows that $\dot{m}(.,\theta,v)$ is a Euclidian class of functions indexed on $N(\theta_0)\times[-\infty,\infty]^k$ with envelope $(B(z_t)\left(2\sqrt{d}\sup_{N(\theta_0)}\|\theta-\theta'\|\right)^{\alpha} + |\partial p_i(z_t|\theta_0)/\partial\theta|)M(\chi_t)$. Then

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

since $\sup_{\|\theta-\theta_0\|\leq\delta}\sup_v \left\|\frac{1}{n}\sum_t^n \left[\dot{m}(z_t\theta,v) - \dot{m}(z_t,\theta_0,v)\right]\right\| = o_p(1)$ by Lemma 2.1 of Arcones and Yu (1994). This completes the proof of 17.

The second part of the result follows from the fact that the class of functions $\mathcal{F} = m_v(.) + \dot{m}(\theta,v)\,l(.,.,\theta_0)$ is a Euclidian 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 variable 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 7. Convergence to a weak limit in $\mathfrak{D}[-\infty,\infty]^k$ then follows again by Lemma 2.1 of Arcones and Yu (1994). ■

We now establish that the process $T\hat{V}(v)$, defined in (7) 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 $-\infty = \lambda_0 < \lambda_1 < \ldots < \lambda_N = \infty$ on $[-\infty,\infty]$, let $\Delta\pi_{\lambda_i} = \pi_{\lambda_{i+1}} - \pi_{\lambda_i}$ and set

$$c_N(V) = \sum_{i=1}^N \left\langle \phi(.,v), \Delta\pi_{\lambda_i}\bar{l}(.,\theta)\right\rangle C_{\lambda_i}^{-1} V(\pi_{\lambda_i}^{\perp}\bar{l}(\vartheta,\theta)). \tag{18}$$

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 $\max_i(\Phi(\lambda_{i+1}) - \Phi(\lambda_i)) \to 0$. Let the limit of $c_N(V)$ be denoted as $c(V) = \int \left\langle \phi(.,v), d\pi_{\lambda}\bar{l}(.,\theta)\right\rangle C_{\lambda}^{-1} V\left(\pi_{\lambda}^{\perp}\bar{l}(.,\theta)\right)$

Condition 8 *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 $\lambda' \downarrow \lambda$. 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 $\pi_\lambda f(v) = \mathbf{1}(v \in A_\lambda) f(v)$ and $\pi_\lambda^\perp = 1 - \pi_\lambda$ such that $\pi_\lambda^\perp f(v) = \mathbf{1}(v \notin A_\lambda) f(v)$. We then define the inner product $\langle f(.), g(.)\rangle := \int f(u)g(u)'dH(u)$ and the matrix*

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

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 8 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 18. 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 7. 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 $[-\infty, \infty]^k$ 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 $\alpha \in \mathbb{R}^k$ 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 $\rho_1(\xi) = |\xi_1| + ... + |\xi_k|$ for any vector $\xi = (\xi_1, ..., \xi_k) \in \mathbb{R}^k$ and $\rho_\infty(\xi) = \max_i |\xi_i|$. Set $\xi = \left\langle \phi, \Delta \pi_\mu \bar{l}(., \theta) \right\rangle$ 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 8 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 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

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

Note that $dH(u)=\left(p(u_2)-p(u_2)^2\right)dF_u\left(u\right)$ such that

$$\begin{aligned}\int\phi\left(u,\tau\right)\bar{l}(u,\theta_0)dH(u) &= \int\phi\left(u,\tau\right)\frac{1}{(p(u_2)-p(u_2)^2)}\frac{\partial p(u_2,\theta_0)}{\partial\theta}dH(u)\\ &= \int\phi\left(u,\tau\right)\frac{\partial p(u_2,\theta_0)}{\partial\theta}dF_u(u)=\dot{m}(\tau,\theta_0)\end{aligned}$$

and

$$\begin{aligned}&\int\bar{l}(u,\theta_0)\bar{l}(u,\theta_0)'dH(u)\\ =&\int\frac{1}{\left(p(u_2)-p(u_2)^2\right)^2}\frac{\partial p(u_2,\theta_0)}{\partial\theta}\frac{\partial p(u_2,\theta_0)}{\partial\theta'}dH(u)\\ =&\int\frac{1}{\left(p(u_2)-p(u_2)^2\right)^2}\frac{\partial p(u_2,\theta_0)}{\partial\theta}\frac{\partial p(u_2,\theta_0)}{\partial\theta'}dF_u(u)=\Sigma_\theta\end{aligned}$$

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

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

$$\begin{aligned}\left\langle\phi(.,v),\bar{l}(.,\theta)\right\rangle &= \int\phi\left(u,v\right)\frac{1}{(p(u_2)-p(u_2)^2)}\frac{\partial p(u_2,\theta_0)}{\partial\theta}dH(u)\\ &= \dot{m}(v,\theta_0)\end{aligned}$$

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 $T\hat{V}(v)=\hat{V}\left(v\right)-\int\left\langle\phi\left(.,v\right),d\pi_\lambda\bar{l}(.,\theta)\right\rangle C_\lambda^{-1}\hat{V}(\pi_\lambda^\perp\bar{l}(.,\theta))$ has covariance function $\Gamma(v,\tau)$ we first consider $E\left(TV(v)\right)^2$ where

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

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

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

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 $\left[-\infty,\infty\right]^k$ 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 2, 3, 4, 5, 6, 7 and 8 are satisfied. Fix* $x < \infty$ *arbitrary and define* $\Upsilon_x = \left\{v \in \left[-\infty,\infty\right]^k | v = \pi_x v\right\}$. *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* $\mathfrak{D}\left[\Upsilon_x\right]$ *where* $\Rightarrow$ *denotes weak convergence.*

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

$$\begin{aligned}
& T\hat{V}_n - TV_n \qquad (19) \\
= \; & -\dot{m}(v,\theta_0) n^{-1/2} \sum_{t=1}^n l\left(D_t, z_t, \theta_0\right) \\
& - \int \left\langle \phi\left(.,v\right), d\pi_\lambda \bar{l}(.,\theta_0)\right\rangle C_\lambda^{-1} \left(\hat{V}_n\left(\pi_\lambda^\perp \bar{l}(.,\theta_0)\right) - V_n\left(\pi_\lambda^\perp \bar{l}(.,\theta_0)\right)\right) + o_p\left(1\right)
\end{aligned}$$

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

$$\begin{aligned}
& \hat{V}_n\left(\pi_\lambda^\perp \bar{l}(.,\theta_0)\right) - V_n\left(\pi_\lambda^\perp \bar{l}(.,\theta_0)\right) \\
= & \; n^{-1/2}\sum_{t=1}^n \mathbf{1}\left\{U_t \notin A_\lambda\right\} \bar{l}(z_t,\theta_0)\left(p(z_t,\theta_0) - p(z_t,\hat{\theta})\right) \\
= & \; n^{-1/2}\sum_{t=1}^n \mathbf{1}\left\{U_t \notin A_\lambda\right\} \bar{l}(z_t,\theta_0)\left(\frac{\partial p(z_t,\theta_n)}{\partial\theta} - \frac{\partial p(z_t,\theta_0)}{\partial\theta}\right)\left(\hat{\theta} - \theta_0\right) \\
& + n^{-1/2}\sum_{t=1}^n \mathbf{1}\left\{U_t \notin A_\lambda\right\} \bar{l}(z_t,\theta_0)\frac{\partial p(z_t,\theta_0)}{\partial\theta}\left(\hat{\theta} - \theta_0\right) \\
: & = R_1(\lambda) + R_2(\lambda).
\end{aligned}$$

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

$$\begin{aligned}
\sup_\lambda \|R_1(\lambda)\| & \leq n^{1/2}\left\|\hat{\theta} - \theta_0\right\| n^{-1}\sum_{t=1}^n \left\|\bar{l}(z_t,\theta_0)\right\| \left\|\frac{\partial p(z_t,\theta_n)}{\partial\theta} - \frac{\partial p(z_t,\theta_0)}{\partial\theta}\right\| \\
& \leq n^{1/2}\left\|\hat{\theta} - \theta_0\right\| n^{-1}\sum_{t=1}^n \left\|\bar{l}(z_t,\theta_0)\right\| \left\|\dot{m}(z_t,\theta_n) - \dot{m}(z_t,\theta_0)\right\| \\
& \quad + n^{1/2}\left\|\hat{\theta} - \theta_0\right\| n^{-1}\sum_{t=1}^n \left\|\bar{l}(z_t,\theta_0)\right\| \left\|\dot{m}(\theta_n) - \dot{m}(\theta_0)\right\| \\
& \leq n^{1/2}\left\|\hat{\theta} - \theta_0\right\| \left(n^{-1}\sum_{t=1}^n \left\|\bar{l}(z_t,\theta_0)\right\|^2\right)^{1/2} \left(n^{-1}\sum_{t=1}^n \left\|\dot{m}(z_t,\theta_n) - \dot{m}(z_t,\theta_0)\right\|^2\right)^{1/2}
\end{aligned}$$

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 $E\left\|\bar{l}((y_t,z_t),\theta_0)\right\| < \infty$ and Lemma 2.1 of Arcones and Yu (1994) this implies that

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

By Condition 6 it follows that

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

for some $\alpha > 0$ such that

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

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

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

Next consider $R_2\left(\lambda\right) - \int \pi_\lambda^\perp \bar{l}(\vartheta,\theta_0)\dot{m}(d\vartheta,\theta_0) n^{1/2}\left(\hat{\theta}-\theta_0\right)$. Note that

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

and

$$\begin{aligned}
& \sup_\lambda \left\| \mathbf{1}\left\{U_t \notin A_\lambda\right\} \bar{l}(z_t,\theta_0)\frac{\partial p(z_t,\theta_0)}{\partial \theta'} \right\| \\
\leq & \left\| \bar{l}(z_t,\theta_0)\frac{\partial p(z_t,\theta_0)}{\partial \theta'} \right\| \\
= & \left\| \frac{\partial p(z_t,\theta_0)}{\partial \theta}\frac{\partial p(z_t,\theta_0)}{\partial \theta'} \frac{1}{p(z_t,\theta_0)\left(1-p(z_t,\theta_0)\right)} \right\| \\
\leq & \left\| \frac{\partial p(z_t,\theta_0)}{\partial \theta} \frac{1}{\left[p(z_t,\theta_0)\left(1-p(z_t,\theta_0)\right)\right]^{1/2}} \right\|^2 \\
= & \sum_{i=1}^d \left(\frac{\partial p_i(z_t,\theta_0)}{\partial \theta}\right)^2 \frac{1}{\left[p(z_t,\theta_0)\left(1-p(z_t,\theta_0)\right)\right]} \\
\leq & d^{-\left(1-\frac{2}{2+\delta}\right)} \left(\sum_{i=1}^d \left(\frac{\partial p_i(z_t,\theta_0)}{\partial \theta}\right)^{4+\delta} \frac{1}{\left[p(z_t,\theta_0)\left(1-p(z_t,\theta_0)\right)\right]^{(4+\delta)/2}} \right)^{2/(4+\delta)}
\end{aligned}$$

with

$$E\left[\sum_{i=1}^d \left(\frac{\partial p_i(z_t,\theta_0)}{\partial \theta}\right)^{4+\delta} \frac{1}{\left[p(z_t,\theta_0)\left(1-p(z_t,\theta_0)\right)\right]^{(4+\delta)/2}}\right] < \infty$$

by Condition 6. This shows that $\left(1-\mathbf{1}\left\{(y_t,z_t)\in A_\lambda\right\}\right)\bar{l}(z_t,\theta_0)\frac{\partial p(z_t,\theta_0)}{\partial \theta'}$ is a Euclidian 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

$$\sup_\lambda \left\| R_2\left(\lambda\right) - \int \pi_\lambda^\perp \bar{l}(\vartheta,\theta_0)\dot{m}(d\vartheta,\theta_0) n^{-1/2}\left(\hat{\theta}-\theta_0\right) \right\| = o_p(1).$$

It then follows that uniformly on Υ_x

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

Now note that $\int \pi_\lambda^\perp \bar{l}(\vartheta,\theta_0)\dot{m}(d\vartheta,\theta_0) = C_\lambda$ such that

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

Substituting back in 19 then shows that $\sup_{v\in\Upsilon_x}\left|T\hat{V}_n(v)-TV_n(v)\right|=o_p(1)$.

For the second part of the proposition consider

$$TV_n(v)=V_n(v)-\int\left\langle\phi(.,v),d\pi_\lambda\bar{l}(.,\theta_0)\right\rangle C_\lambda^{-1}n^{-1/2}\sum_{t=1}^n\mathbf{1}\{U_t\notin A_\lambda\}\bar{l}(z_t,\theta_0)(D_t-p(z_t,\theta_0)).$$

Under H_0 it follows that

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

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

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

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

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

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

$$EY_{1t}^2=\Gamma(v,v),$$

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

and

$$\begin{aligned}EY_{1t}(v)Y_{2t}(v)&=\int\left\langle\phi(.,v),d\pi_\lambda\bar{l}(.,\theta_0)\right\rangle C_\lambda^{-1}E\left[\mathbf{1}\{U_t\notin A_\lambda\}\phi(U_t,v)\partial\log p_t(z_t,\theta_0)/\partial\theta\right]\\&=\int\left\langle\phi(.,v),d\pi_\lambda\bar{l}(.,\theta)\right\rangle C_\lambda^{-1}\left\langle\phi(U_t,v),\pi_\lambda^\perp\bar{l}(.,\theta)\right\rangle\end{aligned}$$

which shows that $EY_t(v)^2=\Gamma(v,v)$. Also, $EY_{1t}(v)Y_{1t}(\tau)=\Gamma(v,\tau)$,

$$EY_{2t}(v)Y_{2t}(\tau)=2\int\left\langle\phi(.,v),d\pi_\lambda\bar{l}(.,\theta)\right\rangle C_\lambda^{-1}\left\langle\phi(.,\tau),\pi_\lambda^\perp\bar{l}(.,\theta)\right\rangle$$

and

$$EY_{1t}(v)\, Y_{2t}(\tau) = \int \left\langle \mathbf{1}\left(. \leq v\right), d\pi_\lambda \bar{l}(.,\theta)\right\rangle C_\lambda^{-1} \left\langle \mathbf{1}\left(. \leq \tau\right), \pi_\lambda^{\perp} \bar{l}(.,\theta)\right\rangle$$

such that $EY_t(v)\, Y_t(\tau) = \Gamma(v,\tau)$. It also follows that $EY_t^2 < \infty$ 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 $\left\|\bar{l}(z_t,\theta_0)\right\|^2 = \sum_{i=1}^d \left|\bar{l}_i(z_t,\theta_0)\right|^2$ such that by the Hölder inequality $\left\|\bar{l}(z_t,\theta_0)\right\|^{2+\delta} \leq d^{\delta/2} \sum_{i=1}^d \left|\bar{l}_i(z_t,\theta_0)\right|^{2+\delta}$ where

$$\left|\bar{l}_i(z_t,\theta_0)\right| \leq \left|\partial p_i(z_t,\theta_0)/\partial\theta\right| \left|\left(p(z_t,\theta_0)\left(1 - p(z_t,\theta_0)\right)\right)^{-1}\right|.$$

By the Cauchy Schwartz inequality it then follows that

$$E\left|\bar{l}_i(z_t,\theta_0)\right|^{2+\delta} \leq \left(E\left|\partial p_i(z_t,\theta_0)/\partial\theta\right|^{4+2\delta}\right)^{1/2} \left(E\left|\left(p(z_t,\theta_0)\left(1 - p(z_t,\theta_0)\right)\right)^{-1}\right|^{4+2\delta}\right) < \infty$$

which is bounded for some δ by Condition (6). This shows that $g(y_t, z_t, v)$ is a Euclidian class of functions and by Lemma 2.14 of Pakes and Pollard it follows that $Y_t(v)$ is a Euclidian 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 2, 3, 4, 5,6, 7 and 8 are satisfied. Fix $x < \infty$ arbitrary and define $\Upsilon_x = \left\{ v \in [-\infty, \infty]^k \,|v = \pi_x v \right\}$. Then,*

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

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

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

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

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

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

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

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 8. 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

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

where $\bar{l}_i(z_t,\theta)$ has the integrable Envelope $B(z_t)\left(2\sqrt{d}\sup_{N(\theta_0)} \|\theta - \theta'\|\right)^\alpha + \left|\bar{l}_i(z_t,\theta_0)\right|$ on $N(\theta_0)$ by Condition 6. By Condition 8 the functions $\mathbf{1}\{(y_t,z_t) \in A_\lambda\}$ form a Euclidian 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

$$\limsup_n P\left(\sup_{\lambda,\theta \in \Upsilon_x \times N(\theta_0)} \sup_{\lambda',\theta' : d((\lambda,\theta),(\lambda',\theta')) < \delta} \left| v_n(\theta',\lambda') - v_n(\theta,\lambda) \right| > \varepsilon \right) < \eta. \tag{20}$$

Property 20 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

$$P\left(\left\| \hat{C}_\lambda - C_\lambda\left(\hat{\theta}\right) \right\| > \varepsilon \right) \leq P\left(\sup_{\lambda,\theta \in \Upsilon_x \times N(\theta_0)} |v_n(\theta,\lambda)| > \varepsilon \right) + P\left(\hat{\theta} \notin N(\theta_0) \right) \xrightarrow{p} 0 \tag{21}$$

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

Then

$$\begin{aligned}
T_n\hat{V}_n(v) - TV_n(v) &= -\dot{m}(v,\theta_0)n^{-1/2}\sum_{t=1}^n l(D_t,z_t,\theta_0) + o_p(1) \\
&\quad - \int d\left(\int \phi(u,v)\, \pi_\lambda \bar{l}(.,\hat{\theta}) d\hat{H}_n(u) \right) \hat{C}_\lambda^{-1} \hat{V}_n(\pi_\lambda^\perp \bar{l}(.,\hat{\theta})) \\
&\quad + \int \left\langle \phi(.,v), d\pi_\lambda \bar{l}(.,\theta_0) \right\rangle C_\lambda^{-1} V_n(\pi_\lambda^\perp \bar{l}(.,\theta_0)).
\end{aligned}$$

>From before we have

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

where

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

by 21. Next we consider

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

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

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

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

$$\begin{aligned}
& E\left[\mathbf{1}\left\{U_t\notin A_\lambda\right\}\frac{\partial\bar{l}(\left(y_t,z_t\right),\theta_0)}{\partial\theta'}\left(D_t-p(z_t,\theta_0)\right)|z_t\right] \\
= & E\left[\mathbf{1}\left\{U_t\notin A_\lambda\right\}\frac{\partial\bar{l}(\left(y_t,z_t\right),\theta_0)}{\partial\theta'}|z_t\right]E\left[\left(D_t-p(z_t,\theta_0)\right)|z_t\right]=0
\end{aligned}$$

under H_0 such that finite dimensional convergence follows by the martingale difference CLT and uniform convergence follows from the fact that $\mathbf{1}\left\{U_t \notin A_\lambda\right\} \frac{\partial \bar{l}(U_t,\theta_0)}{\partial \theta'} \left(D_t - p(z_t,\theta_0)\right)$ is a Euclidian class of functions by Condition 8. 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 $R_{22}\left(\lambda, \theta_n\right)$ we note that

$$E\left[\mathbf{1}\left\{U_t \notin A_\lambda\right\} \frac{\partial \bar{l}(U_t,\theta)}{\partial \theta'} \left(D_t - p(z_t,\theta_0)\right) |z_t\right] = 0$$

for any θ. By Lemma 2.1 of Arcones and Yu it thus follows that $R_{22}\left(\lambda, \theta\right)$ 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 $\limsup P\left[\sup_{\theta: d(\theta,\theta_0)\leq\delta} \left|R_{22}\left(\lambda,\theta\right)\right| > \varepsilon\right] < \eta$ for all $\varepsilon, \eta > 0$ and some $\delta > 0$. Use root-n convergence of θ_n to conclude from this that $R_{22}\left(\lambda, \theta_n\right) = o_p(1)$. 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 $R_4\left(\lambda\right)$ where

$$R_4\left(\lambda\right) = n^{-1}\sum_{t=1}^{n} \mathbf{1}\left\{U_t \notin A_\lambda\right\} \frac{\partial \bar{l}(U_t,\theta_0)}{\partial \theta} \frac{\partial p(z_t,\theta_0)}{\partial \theta'}$$

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

For $R_3\left(\lambda\right)$ note that

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

uniformly converges to

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

We have thus established that

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

Using this result we obtain

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

The leading term is then

$$\begin{aligned}\int d\left(\int \phi\left(u,v\right)\pi_\lambda \bar{l}(u,\hat{\theta})d\hat{H}_n(u)\right) &= \int d\left(\int \phi\left(u,v\right)\pi_\lambda \bar{l}(u,\theta_0)dH_n(u)\right) \\ &\quad + \int d\left(\int \phi\left(u,v\right)\pi_\lambda \frac{\partial^2 p(z_t,\theta_n)}{\partial\theta\partial\theta'}d\hat{F}_u(u)\right)\left(\hat{\theta} - \theta_0\right)\end{aligned} \qquad (22)$$

where $\hat{F}_u(u)$ is defined in (24) in Appendix B.1 and

$$\begin{aligned}\left\| \int d \int \phi\left(u,v\right) \pi_\lambda \frac{\partial^2 p(z_t,\theta_n)}{\partial\theta\partial\theta'} d\hat{F}_u(u) \right\| &\leq n^{-1} \sum_{t=1}^{n} \left\| \mathbf{1}\left\{U_t \leq v\right\} \mathbf{1}\left\{U_t \in A_\lambda\right\} \frac{\partial^2 p(z_t,\theta_n)}{\partial\theta\partial\theta'} \right\| \\ &\leq n^{-1} \sum_{t=1}^{n} \left\| \frac{\partial^2 p(z_t,\theta_0)}{\partial\theta\partial\theta'} \right\| + n^{-1} \sum_{t=1}^{n} \left\| \frac{\partial^2 p(z_t,\theta_n)}{\partial\theta\partial\theta'} - \frac{\partial^2 p(z_t,\theta_0)}{\partial\theta\partial\theta'} \right\| \\ &\leq n^{-1} \sum_{t=1}^{n} \left\| \frac{\partial^2 p(z_t,\theta_0)}{\partial\theta\partial\theta'} \right\| + C \left\|\theta_n - \theta_0\right\|^\alpha n^{-1} \sum_{t=1}^{n} B(z_t) = O_p(1)\end{aligned}$$

where C is a finite constant, the third inequality uses Condition (6) and the last equality follows from a standard law of large numbers for strong mixing sequences. The first term in 22 then is

$$\int d \left(\int \phi\left(u,v\right) \pi_\lambda \bar{l}(u,\theta_0) dH_n(u) \right) = n^{-1} \sum_{t=1}^{n} \phi\left(U_t,v\right) \mathbf{1}\left\{U_t \in A_\lambda\right\} \frac{\partial p(z_t,\theta_0)}{\partial\theta}$$

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

Finally we need to show that

$$\int \left(d\left(\int \phi\left(u,v\right) \pi_\lambda \bar{l}(u,\theta_0) dH_n(u) \right) - \left\langle \phi\left(.,v\right), d\pi_\lambda \bar{l}(.,\theta_0) \right\rangle \right) C_\lambda^{-1} V_n(\pi_\lambda^\perp \bar{l}(u,\theta_0)) = o_p(1). \tag{23}$$

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 $v \in \Upsilon_x$,

$$\int \phi\left(.,v\right) \pi_\lambda \bar{l}(.,\theta_0) dH_n(v) - \left\langle \phi\left(.,v\right), \pi_\lambda \bar{l}(.,\theta_0) \right\rangle = n^{-1} \sum_{t=1}^{n} g(z_t,\lambda,v) - E\left(g(z_t,\lambda,v)\right) \to 0 \text{ a.s.}$$

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 23 can now be written as $\int_{-\infty}^{x} \zeta_n\left(\lambda\right)'\left(G_n(d\lambda) - G(d\lambda)\right)$ such that for any $\delta > 0$

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

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

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

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

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

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

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

which implies that for some N_0 and all $n > N_0$, $\left| \int_{-\infty}^{x} \zeta_i(\lambda) \left(G_{ni}(d\lambda, v) - G_i(d\lambda, v)\right) \right| < 3\varepsilon$ uniformly on $H \times \Upsilon_x$ by the arguments of Chang (1994, p.396) which establishes 23. 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

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

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

Condition 9 *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 $\Psi(r) = (2\pi)^k \int \exp(ir'x)K(x)dx$ that satisfies $\int |\Psi(r)|\, dr < \infty$ where $i = \sqrt{-1}$.

Theorem 7 *Assume Conditions 2, 3, 4, 5,6, 7, 8 and 9 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 Euclidian class with envelope M and a measurable map with envelope M_1 forms another Euclidian class with envelope $M \circ M_1$. Since $F_k(x_k|x_{k-1}, ..., x_1)$ is 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 Euclidian 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

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

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.

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

$$\hat{F}_u(v) = n^{-1} \sum_{t=1}^{n} \{U_t \leq v\}, \tag{24}$$

and let

$$\begin{aligned} H_n(v) &= \int_{-\infty}^{v} \left(p(u_2, \theta_0) - p(u_2, \theta_0)^2\right) d\hat{F}_u(u) \\ &= n^{-1} \sum_{t=1}^{n} \left(p(z_t, \theta_0) - p(z_t, \theta_0)^2\right) \mathbf{1}\{U_t \leq v\} \end{aligned}$$

as well as

$$\begin{aligned} \hat{H}_n(v) &= \int_{-\infty}^{v} \left(p(u_2, \hat{\theta}) - p(u_2, \hat{\theta})^2\right) d\hat{F}_u(u) \\ &= n^{-1} \sum_{t=1}^{n} \left(p(z_t, \hat{\theta}) - p(z_t, \hat{\theta})^2\right) \mathbf{1}\{U_t \leq v\}. \end{aligned}$$

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

$$\begin{aligned} \hat{C}_\lambda &= \int \pi_\lambda^\perp \bar{l}(v, \hat{\theta}) \pi_\lambda^\perp \bar{l}(v, \hat{\theta})' d\hat{H}_n(v) \\ &= n^{-1} \sum_{t=1}^{n} \left(1 - \mathbf{1}\{U_t \in A_\lambda\}\right) \bar{l}(U_t, \hat{\theta}) \bar{l}(U_t, \hat{\theta})' \left(p(z_t, \hat{\theta}) - p(z_t, \hat{\theta})^2\right) \end{aligned}$$

such that

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

where

$$\int \phi\{u, v\} \pi_\lambda \bar{l}(., \hat{\theta}) d\hat{H}_n(u) = n^{-1} \sum_{t=1}^{n} \phi(U_t, v) \mathbf{1}\{U_t \in A_\lambda\} \frac{\partial p(z_t, \hat{\theta})}{\partial \theta}.$$

Finally, write

$$\hat{V}_n(\pi_\lambda^\perp \bar{l}(u, \hat{\theta})) = n^{-1/2} \sum_{t=1}^{n} \left(1 - \mathbf{1}\{U_t \in A_\lambda\}\right) \bar{l}(U_t, \hat{\theta}) \left(D_t - p(z_t, \hat{\theta})\right).$$

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

$$\hat{C}_\lambda = n^{-1} \sum_{t=1}^{n} \mathbf{1}\left\{y_{1t} > \lambda\right\} \bar{l}(z_t, \hat{\theta}) \bar{l}(z_t, \hat{\theta})' \left(p(z_t, \hat{\theta}) - p(z_t, \hat{\theta})^2 \right), \tag{25}$$

$$\hat{V}_n(\pi_\lambda^\perp \bar{l}(u, \hat{\theta})) = n^{-1/2} \sum_{t=1}^{n} \mathbf{1}\left\{y_{1t} > \lambda\right\} \bar{l}(U_t, \hat{\theta}) \left(D_t - p(z_t, \hat{\theta}) \right) \tag{26}$$

and

$$\int \phi(u, v) \pi_\lambda \bar{l}(u, \hat{\theta}) d\hat{H}_n(u) = n^{-1} \sum_{t=1}^{n} \phi\left\{U_t, v\right\} \mathbf{1}\left\{y_{1t} \leq \lambda\right\} \frac{\partial p(z_t, \theta)}{\partial \theta} \tag{27}$$

Combining 25, 26 and 27 then leads to the formulation 9.

B.2 Details for the Rosenblatt Transform

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

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

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

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

It therefore follows that $\hat{V}_{w,n}\left(v\right)$ can be approximated by $V_{w,n}\left(v\right) - \dot{m}_w(w, \theta_0)' n^{-1/2} \sum_{t=1}^{n} l\left(D_t, z_t, \theta_0\right)$. This approximation converges to a limiting process $\hat{V}_w\left(v\right)$ with covariance function

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

where

$$\Gamma_w\left(w, \tau\right) = \int_{[0,1]^k} \phi(u, w) \phi(u, \tau) \left(p(\left[T_R^{-1}\left(u\right)\right]_z) - p(\left[T_R^{-1}\left(u\right)\right]_z)^2 \right) du.$$

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

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

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 $A_{w,0} = \varnothing$, $A_{w,1} = [0,1]^k$, $\lambda \leq \lambda' \implies A_{w,\lambda} \subset A_{w,\lambda'}$ and $A_{w,\lambda'} \backslash A_{w,\lambda} \to \varnothing$ as $\lambda' \downarrow \lambda$. We then define the inner product $\langle f(.), g(.)\rangle_w := \int_{[0,1]^k} f(w)g(w)' dH_w(w)$ where

$$H_w(w) = \int_{u \leq w} \left(p(\left[T_R^{-1}(u)\right]_z, \theta) - p(\left[T_R^{-1}(u)\right]_z, \theta)^2\right) du$$

and the matrix

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

and define the transform $T_w V_w(w)$ 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

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

with $h_w(.) = p(\left[T_R^{-1}(.)\right]_z, \theta)\left(1 - p(\left[T_R^{-1}(.)\right]_z, \theta)\right)$ and $B_w(w)$ 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

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

where $\hat{C}_{w_{s1}} = n^{-1} \sum_{t=1}^{n} \mathbf{1}\{w_{t1} > w_{s1}\} \bar{l}(z_t, \hat{\theta}) \bar{l}(z_t, \hat{\theta})' \left(p(z_t, \hat{\theta}) - p(z_t, \hat{\theta})^2\right)$.

References

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 (2004): "Measuing the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *NBER Working Paper 10220.*

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.

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.

Chesher, A., and I. Jewitt (1987): "The Bias of a Heteroscedasticity-Consistent Covariance Matrix Estimator," *Econometrica*, 55, 1217–1222.

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.

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.

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.

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.

Heckman, J. J., H. Ichimura, and P. E. Todd (1998): "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65, 261–294.

Hirano, K., and G. Imbens (2004): "The Propensity Score with Continuous Treatments," *Berkeley Department of Economics, mimeo.*

Hirano, K., G. Imbens, and G. Ridder (2003): "Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score," *Econometrica*, 71, 1161–1189.

Justel, A., D. Pena, and R. Zamar (1997): "A multivariate Kolmogorov-Smirnov test of goodness of fit," *Statistics and Probability Letters*, 35, 251–259.

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 (2003): "Inference of the Quantile Regression Process," *Forthcoming Econometrica.*

Koul, H. L., and W. Stute (1999): "Nonparametric Model Checks for Time Series," *Annals of Statistics*, pp. 204–236.

Leeper, E. M. (1997): "Narrative and VAR approaches to Monetary Policy: Common Identification Problems," *Journal of Monetary Economics*, pp. 641–657.

Linton, O., and P. Gozalo (1999): "Conditional Independence Restrictions: Testing and Estimation," *mimeo.*

Lucas, R. E. (1972): "Expectations and the Neutrality of Money," *Journal of Economic Theory*, pp. 103–124.

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

Pollard, D. (1984): *Convergence of Stochastic Processes*. Springer-Verlag New York, Inc.

Robins, J. M., S. Greenland, and F.-C. Hu (1999): "Estimation of the Causal Effect of a Time-Varying Exposure on the Marginal Mean of a Repeated Binary Outcome," *Journal of the American Statistical Association*, pp. 687–712.

Robins, J. M., S. D. Mark, and W. K. Newey (1992): "Estimating Exposure Effects by Modelling the Expectation of Exposure Conditional on Confounders," *Biometrics*, pp. 479–495.

Romer, C. D., and D. H. Romer (1989): "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz," *NBER Macroeconomics Annual*, pp. 121–170.

——— (1994): "Monetary Policy Matters," *Journal of Monetary Economics*, pp. 75–88.

——— (1997): "Identification and the Narrative Approach: A Reply to Leeper," *Journal of Monetary Economics*, 40, 659–665.

——— (2004): "A New Measure of Monetary Shocks: Derivation and Implications," *The American Economic Review*, 94, 1055–1084.

Rosenbaum, P., and D. B. Rubin (1985): "Constructing a Control Group using Multivariate Matching Methods that include the Propensity Score," *American Statistician*, 39, 33–38.

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.

Rubin, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies," *Journal of Educational Psychology*, pp. 688–701.

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. A. (1972): "Money, Income and Causality," *American Economic Review*, pp. 540–562.

——— (1980): "Macroeconomics and Reality," *Econometrica*, pp. 1–48.

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.

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.

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

		Rejection Probabilities					
		VM-MC	md$_a$	md$_b$	d$_1$	d$_2$	t-test
γ	β	(1)	(2)	(3)	(4)	(5)	(6)
		A. Sample Size = 100					
0	-0.5	0.096	0.070	0.036	0.070	0.042	0.072
0.5	-0.5	0.140	0.148	0.064	0.080	0.170	0.178
1	-0.5	0.394	0.468	0.292	0.226	0.496	0.574
2	-0.5	0.810	0.888	0.780	0.456	0.906	0.960
0	0	0.082	0.064	0.026	0.046	0.056	0.050
0.5	0	0.154	0.162	0.070	0.068	0.182	0.188
1	0	0.438	0.500	0.328	0.298	0.506	0.570
2	0	0.814	0.906	0.834	0.612	0.862	0.952
0	0.5	0.098	0.060	0.030	0.042	0.060	0.048
0.5	0.5	0.264	0.188	0.088	0.096	0.194	0.202
1	0.5	0.548	0.534	0.360	0.406	0.486	0.616
2	0.5	0.872	0.930	0.868	0.840	0.822	0.970
0	0.9	0.210	0.064	0.010	0.040	0.060	0.042
0.5	0.9	0.436	0.252	0.122	0.180	0.200	0.276
1	0.9	0.766	0.744	0.606	0.616	0.664	0.804
2	0.9	0.928	0.252	0.186	0.158	0.244	0.402
		B. Sample Size = 200					
0	-0.5	0.096	0.058	0.018	0.064	0.054	0.052
0	0	0.084	0.072	0.020	0.052	0.080	0.058
0	0.5	0.104	0.066	0.024	0.050	0.066	0.078
0	0.9	0.226	0.044	0.012	0.034	0.050	0.062

Table 1: Rejection Probabilities from a dynamic Logit Model

	k=2		k=3		k=4	
	md	d	md	d	md	d
1-α	(1)	(2)	(3)	(4)	(5)	(6)
0.5	0.17555	0.13877	0.1224	0.079614	0.08127	0.045061
0.8	0.36124	0.29359	0.21503	0.14446	0.12871	0.073065
0.9	0.51805	0.43536	0.28873	0.20363	0.16503	0.097858
0.95	0.68209	0.58862	0.36511	0.26808	0.20114	0.12482
0.975	0.85668	0.7454	0.44198	0.33422	0.23826	0.15462
0.99	1.081	0.96801	0.5486	0.42748	0.28919	0.19535
0.995	1.2597	1.1296	0.62995	0.4994	0.32922	0.22667
0.999	1.6911	1.573	0.8238	0.68994	0.4225	0.30895
0.9995	1.9174	1.7816	0.91185	0.77078	0.46407	0.33938
0.9999	2.2286	2.1684	1.083	0.99037	0.53436	0.40949

Table 2: Critical Values based on 100,000 Simulation Replications

	Control variables (lagged)					
	output		output		output	
			inflation		inflation	
					unemployment	
	estimate	p-value	estimate	p-value	estimate	p-value
Lagged Romer Dummies	(1)	(2)	(3)	(4)	(5)	(6)
RD(-1)	0.0129	0.093	0.0125	0.057	0.0150	0.035
	(0.0076)		(0.0065)		(0.0070)	
RD(-2)	-0.0218	0.037	-0.0210	0.022	-0.0176	0.063
	(0.0104)		(0.0091)		(0.0094)	
RD(-3)	-0.0176	0.123	-0.0145	0.219	-0.0146	0.159
	(0.0113)		(0.0117)		(0.0103)	
RD(-4)	-0.0089	0.292	-0.0043	0.644	-0.0020	0.801
	(0.0084)		(0.0093)		(0.0079)	
RD(-5)	0.0013	0.895	0.0042	0.724	-0.0001	0.99
	(0.0101)		(0.0119)		(0.0109)	
RD(-6)	-0.0057	0.278	-0.0031	0.543	-0.0078	0.279
	(0.0052)		(0.0051)		(0.0072)	
RD(-7)	-0.0182	0.105	-0.0142	0.214	-0.0118	0.216
	(0.0112)		(0.0114)		(0.0095)	
RD(-8)	-0.0248	0.011	-0.0238	0.029	-0.0143	0.179
	(0.0096)		(0.0108)		(0.0105)	
RD(-9)	-0.0122	0.386	-0.0157	0.235	-0.0122	0.371
	(0.0140)		(0.0131)		(0.0136)	
RD(-10)	-0.0228	0.014	-0.0221	0.02	-0.0235	0.002
	(0.0092)		(0.0094)		(0.0074)	
RD(-11)	-0.0107	0.199	-0.0075	0.336	-0.0060	0.383
	(0.0083)		(0.0078)		(0.0068)	
RD(-12)	0.0019	0.847	0.0035	0.743	0.0056	0.613
	(0.0099)		(0.0106)		(0.0111)	
R2	0.3888		0.4358		0.5243	
F	2.42		2.05		1.63	
(p-val)	(0.0069)		(0.0250)		(0.0908)	
F-robust	2.27		2.0900		1.99	
(p-val)	(0.0115)		(0.0215)		(0.0303)	

Table 3: Granger Causality Tests using Quarterly Data. Models include 8 lags of the control variables indicated in the column headings. Robust standard errors are reported in brakets. The F-statistic is for the joint significance of the lagged Romer Dummies. The robust F-Statistic was computed using White standard errors. The sample includes 160 quarters from 1952-91.

Future Output	md_b			Logit		
Variable	(1)	(2)	(3)	(4)	(5)	(6)
yn(1)	$[0.06, 0.15]$	$[0.06, 0.15]$	$[0.06, 0.15]$	0.0241	0.0777	0.0604
yn(2)	$[0.15, 0.3]$	$[0.6, 1]$	$[0.3, 0.6]$	0.1915	0.1344	0.2113
yn(3)	$[0.003, 0.006]$	$[0.03, 0.06]$	$[0.006, 0.03]$	0.1774	0.0536	0.0494
yn(4)	$[0, 0.0006]$	$[0.15, 0.3]$	$[0.6, 1]$	0.8805	0.4214	0.2928
yn(5)	$[0, 0.0006]$	$[0.3, 0.6]$	$[0.3, 0.6]$	0.0525	0.1572	0.3009
yn(6)	$[0.06, 0.15]$	$[0.6, 1]$	$[0.6, 1]$	0.8819	0.9706	0.7651
yn(7)	$[0.0006, 0.003]$	$[0.3, 0.6]$	$[0.6, 1]$	0.3144	0.2382	0.2135
yn(8)	$[0, 0.0006]$	$[0.0006, 0.003]$	$[0.0006, 0.003]$	0.0227	0.0129	0.0174
Forecasts	full sample	out-of-sample	out-of-sample	full sample	out-of-sample	out-of-sample
Lagged IP controls	No	No	Yes	No	No	Yes

Table 4: P-values for the md-statistic and parametric Logit. Square brakets indicate that actual p-value lies in the interval of values reported in the table. p-values for the md-statistic are based on simulated critical values reported in Table 2 for the d-statistic and are adjusted to provide a bound as described in the main text. In particular, we use critical values for d and k=3 from Table 2. The corresponding significance levels are then 6α. We report a confidence level interval because the quantiles of the distribution need to be computed numerically.

Future Output	md_a		
Variable	(1)	(2)	(3)
yn(1)	[0.025, 0.05]	[0.025, 0.05]	[0.025, 0.05]
yn(2)	[0.1, 0.2]	[0.2, 0.5]	[0.2, 0.5]
yn(3)	[0.001, 0.005]	[0.01, 0.025]	[0.01, 0.025]
yn(4)	[0.0001, 0.0005]	[0.05, 0.1]	[0.2, 0.5]
yn(5)	[0.0001, 0.0005]	[0.1, 0.2]	[0.1, 0.2]
yn(6)	[0.025, 0.05]	[0.2, 0.5]	[0.2, 0.5]
yn(7)	[0.001, 0.005]	[0.1, 0.2]	[0.2, 0.5]
yn(8)	[0, 0.0006]	[0.0005, 0.001]	[0.0005, 0.001]
Forecasts	full sample	out-of-sample	out-of-sample
Lagged IP controls	No	No	Yes

Table 5: P-values for the md-statistic. Square brakets indicate that actual p-value lies in the interval of values reported in the table. p-values for the md-statistic are based on simulated critical values reported in Table 2 for the md-statistic. In particular, we use critical values for md and k=3 from Table 2. We report a confidence level interval because the quantiles of the distribution need to be computed numerically.

Variable	**Definition**
IPN	Industrial Production, total Index not seasonally adjusted, revised 1990
output	Growth Rate Industrial Production New : $\Delta \ln(\mathrm{IPN})$
RD	Original Romer Dummy
CPU	Consumer Price Index, all urban consumers, not seasonally adjusted
inflation	Inflation rate: $\Delta\ln(\mathrm{CPU}_t)$

Table 6: Data Source and Variable Definitions