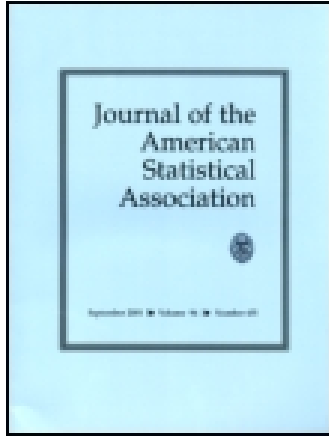


This article was downloaded by: [Dalhousie University]

On: 14 June 2015, At: 09:12

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of the American Statistical Association

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/uasa20>

Quantifying Causal Effects of Road Network Capacity Expansions on Traffic Volume and Density via a Mixed Model Propensity Score Estimator

Daniel J. Graham, Emma J. McCoy & David A. Stephens

Accepted author version posted online: 01 Oct 2014. Published online: 22 Dec 2014.



CrossMark

[Click for updates](#)

To cite this article: Daniel J. Graham, Emma J. McCoy & David A. Stephens (2014) Quantifying Causal Effects of Road Network Capacity Expansions on Traffic Volume and Density via a Mixed Model Propensity Score Estimator, Journal of the American Statistical Association, 109:508, 1440-1449, DOI: [10.1080/01621459.2014.956871](https://doi.org/10.1080/01621459.2014.956871)

To link to this article: <http://dx.doi.org/10.1080/01621459.2014.956871>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Quantifying Causal Effects of Road Network Capacity Expansions on Traffic Volume and Density via a Mixed Model Propensity Score Estimator

Daniel J. GRAHAM, Emma J. MCCOY, and David A. STEPHENS

Road network capacity expansions are frequently proposed as solutions to urban traffic congestion but are controversial because it is thought that they can directly “induce” growth in traffic volumes. This article quantifies causal effects of road network capacity expansions on aggregate urban traffic volume and density in U.S. cities using a mixed model propensity score (PS) estimator. The motivation for this approach is that we seek to estimate a dose-response relationship between capacity and volume but suspect confounding from both observed and unobserved characteristics. Analytical results and simulations show that a longitudinal mixed model PS approach can be used to adjust effectively for time-invariant unobserved confounding via random effects (RE). Our empirical results indicate that network capacity expansions can cause substantial increases in aggregate urban traffic volumes such that even major capacity increases can actually lead to little or no reduction in network traffic densities. This result has important implications for optimal urban transportation strategies. Supplementary materials for this article are available online.

KEY WORDS: Average treatment effect; Confounding; Congestion; Continuous dose-response; Urban traffic volumes.

1. INTRODUCTION

Road network congestion, which is characterized by travel times in excess of those under free-flow conditions, is a phenomenon experienced in most major urban areas around the world. Transport engineers generally view congestion as undesirable because it can impose large costs on road users and the economy through delay and reduced traffic flow. A common intervention used to ameliorate the problem is to build more roads in the hope of reducing traffic densities and accommodating future growth in traffic volumes. This strategy is, however, controversial because some believe that network capacity expansions themselves cause a direct rise in aggregate traffic volumes, a phenomenon known as “induced demand.”

The theory of induced demand is widely known (for key concepts see Small and Verhoef 2007; Kelly 2008). In principle, it can occur via the following causal process. A capacity expansion (i.e., increase in lane miles) is introduced which causes an immediate reduction in traffic density (i.e., traffic volume to capacity ratio) on the network and consequently network congestion falls. With less congestion, average travel times decrease causing the volume of aggregate traffic to rise in response because the *cost* of travel, in terms of time spent, has effectively fallen. As traffic volumes rise, travel times start to increase as congestion again becomes prevalent. In theory these interactions between demand and cost can yield a new equilibrium with a higher volume of network traffic than prior to the capacity expansion. Depending on the extent of induced demand, undesirable consequences of road use such as congestion, pollution, and accidents could worsen. Many U.S. cities have experienced

considerable expansions in road capacity in recent decades and the purpose of this article is to investigate consequences for road traffic volumes and densities.

There are a number of existing empirical studies that have sought statistical evidence for induced demand (for reviews see Noland 2001; Cervero 2002). Cervero (2002) conducted a meta-analysis of such studies and finds that while the statistical work generally finds evidence of induced demand, there is substantial variation in estimates leading to ambiguity over magnitude. This is in part due to contextual differences between studies, for instance in spatial coverage and units used for analysis, but crucially also due to the statistical model specifications adopted.

There are three key issues which have guided model specification: confounding from measured covariates, potential for unmeasured confounding, and bi-directionality between demand and capacity. Some studies have addressed confounding using a linear regression with adjustment for covariates and, where appropriate, dummy variables for unmeasured categorical effects (e.g., Hansen and Huang 1997; Fulton et al. 2000; Cervero 2003). Other studies have explicitly recognized the bi-directional nature of the relationship between travel demand and network capacity assignment either through simultaneous equation modeling of assignment and outcome or through application of instrumental variables (IV) estimation to the linear model (e.g., Noland and Cowart 2000; Noland 2001; Cervero and Hansen 2002; Hymel, Small, and Dender 2010; Duranton and Turner 2011). Finally, there are studies that have adopted Granger causality methods (e.g., Granger 1969) to test whether changes in capacity help to predict future values of traffic volume given the dynamic evolution of volume itself over time (e.g., Fulton et al. 2000; Cervero and Hansen 2002; Melo, Graham, and Canavan 2013).

These previous attempts to obtain estimates of induced demand have used outcome regression (OR) models. In this

Daniel Graham is Professor, Department of Civil & Environmental Engineering, Imperial College London, London, SW7 2AZ, UK (E-mail: d.j.graham@imperial.ac.uk). Emma McCoy is Deputy Head of Department, Department of Mathematics, Imperial College London, SW7 2AZ, UK (E-mail: e.mccoy@imperial.ac.uk). David A. Stephens is Professor, Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Street West, Montreal, Quebec, Canada (E-mail: dstephens@math.mcgill.ca). The authors thank Erica Moodie, Robin Henderson, Nick Heard, the associate editor, and three anonymous referees for helpful comments.

article, we estimate the relationship between network capacity and aggregate traffic volumes and densities in U.S. cities using a novel mixed model generalized propensity score (GPS) estimator. To our knowledge, this is the first time that induced demand has been studied using a statistical model based on the potential outcomes framework for causal inference. The original GPS approach was introduced by Imbens (2000) and Hirano and Imbens (2004) for estimation of multivalued and continuous treatment effects with measured confounding. Here, we provide a longitudinal mixed model extension of the GPS estimator which can accommodate both measured time-varying confounding and unmeasured time-invariant confounding, as well as bidirectionality between response and treatment assignment. This is attractive for the type of application we present for a number of reasons. First, it addresses the three key estimation issues emphasized in previous literature. Second, it allows us to estimate a flexible semiparametric dose-response function rather than a single summary point estimate, which can reveal heterogeneity in treatment effects by dose. Third, the formal framework for causal inference that we adopt requires us to ensure that the sample for causal comparison is valid, that is, that it comprises units that are genuinely comparable and so focuses the inference problem on the relevant subsample and avoids undue extrapolation.

The article is structured as follows. Section 2 explains the properties of mixed model GPS methods for continuous dose-response estimation. Section 3 illustrates some properties of the mixed model PS approach through simulation. Section 4 presents results from our case study application. Section 5 concludes and outlines some issues for future research.

2. A MIXED MODEL PROPENSITY SCORE (PS) APPROACH FOR ESTIMATION OF CONTINUOUS TREATMENT EFFECTS

2.1 The PS for Continuous Treatments in a One-Period Setting

In this article, we are concerned with estimating causal quantities associated with a continuous treatment. We consider first the one-period setting: the observed data are realizations of a random vector, $W_i = (Y_i, D_i, X_i, U_i)$, $i = 1, \dots, N$, where for the i th unit of observation Y_i denotes a response, D_i the treatment received, X_i is a vector of covariates which we assume are sufficient to represent all sources of confounding, and U_i is a vector of covariates that are not confounders because they are independent of Y_i but not D_i . The treatment can take values in some bounded interval in \mathbb{R} , $\mathcal{D} \subseteq \mathbb{R}$, and we define the indicator for receipt of treatment level (or dose) d as

$$I_{D_i}(d) = \begin{cases} 1 & \text{if } D_i = d \\ 0 & \text{if } D_i \neq d \end{cases}$$

For any treatment level $d \in \mathcal{D}$, we assume the existence of a *potential outcome* for unit i , which we denote $Y_i(d)$. The set of all potential outcomes is $\mathcal{Y}_i = \{Y_i(d), d \in \mathcal{D} \text{ for } i = 1, \dots, N\}$ and the full data for estimation of causal effects are taken to be $(\mathcal{Y}_i, D_i, X_i, U_i)$. Unobserved potential outcomes are then treated much like missing data.

For continuous treatments, Hirano and Imbens (2004) showed that observance of actual outcomes is sufficient to estimate av-

erage potential outcomes (APOs) for a population, even under nonrandom treatment assignments, so long as three key assumptions hold. First, to address confounding we must be able to assume “ignorability,” or conditional independence between the response and treatment assignment given the covariates X_i : $Y_i(d) \perp\!\!\!\perp I_{D_i}(d) | X_i$ for all $d \in \mathcal{D}$. Second, to ensure comparability across potential outcomes there must be common support, or overlap, by treatment status in the covariate distributions within some region of dose $\mathcal{C} \subseteq \mathcal{D}$. A sufficient condition is that for any subset of \mathcal{C} , say $\mathcal{A} \subseteq \mathcal{C}$, $\Pr(D_i \in \mathcal{A} | X_i = x) > 0$ for all x . Third, for logical consistency there must be equivalence between the observed response under a given dose d and the potential response under that dose: $Y_i \equiv I_{D_i}(d)Y_i(d)$ for all $d \in \mathcal{D}$, for all $Y_i(d) \in \mathcal{Y}_i$, and for $i = 1, \dots, N$. One important implication of this is that the stable unit treatment value assumption (SUTVA) (e.g., Rubin 1978) must hold, which requires: (i) that the outcome for each unit be independent of the treatment status of other units, or in other words, there should be no interference in treatment effects across units; and (ii) that there are no different versions of the treatment.

If these assumptions hold, then the APO under a given dose $D = d$, which we denote $\mu(d) = \mathbb{E}[Y_i(d)]$, or the dose-response function, can be derived as

$$\begin{aligned} \mathbb{E}[Y_i(d)] &= \mathbb{E}_X [\mathbb{E}(Y_i(d) | X_i)] = \mathbb{E}_X [\mathbb{E}(Y_i(d) | I_{D_i}(d), X_i)] \\ &= \mathbb{E}_X [\mathbb{E}(Y_i | I_{D_i}(d), X_i)], \end{aligned} \quad (1)$$

where the second equality follows from ignorability, the third from the SUTVA, and the overlap assumption ensures that the APO is estimable since there are comparable units across treatment levels. In the literature, the subsample of observations trimmed to ensure overlap in covariate distributions is referred to as the “common support region.” Average treatment effects (ATEs) can then be calculated by comparing the APO at different treatment levels. Note that since observations in the common support region may not be representative of the overall population, and since treatment effects may be heterogeneous across individuals, causal parameters (i.e., ATEs and APOs) estimated under imposition of common support will not in general be equivalent to the corresponding population parameters (see, e.g., Crump et al. 2009).

For continuous treatments Hirano and Imbens (2004) showed that we can use a GPS in place of covariates X_i to adjust for confounding. Let $r(d, x_i) = f_{D|X}(d|x_i)$ denote the conditional density function for receiving a particular level (dose) of the treatment (d) given pretreatment variables $X_i = x_i$. The GPS for dose d is then defined as a random variable, which we denote $R_{d,i} = r(d, X_i)$, a scalar function of X_i for fixed d . We may also define the GPS for observed doses, $R_i = r(D_i, X_i)$, which is evaluated at the level of treatment actually received and is a random variable defined with respect to the joint distribution for observed (D_i, X_i) . Clearly when $D_i = d$, $R_i \equiv R_{d,i}$.

Two key properties of the GPS are (i) that balancing follows directly from its definition: $I_{D_i}(d) \perp\!\!\!\perp X_i | r(d, X_i)$; and (ii) given balancing, ignorability can be established conditional on the scalar GPS rather than the potentially high-dimensional covariate vector X_i : $Y_i(d) \perp\!\!\!\perp I_{D_i}(d) | r(d, X_i)$. Using these two properties Hirano and Imbens (2004) showed that if ignorability holds, and the model is correctly specified, the GPS provides a

bias-removal strategy in the context of continuous treatments:

$$\begin{aligned}\mu(d) &\equiv \mathbb{E}[Y_i(d)] \equiv \mathbb{E}_X[\mathbb{E}(Y_i(d)|X_i)] \\ &= \mathbb{E}_X[\mathbb{E}[Y_i(d)|r(d, X_i)]] = \mathbb{E}_X[\mathbb{E}[Y_i|I_{D_i}(d), r(d, X_i)]].\end{aligned}\quad (2)$$

In practice, the following algorithm can be used to estimate the dose-response curve via the GPS:

1. Estimate a model for the conditional density of D given X with parameter vector θ : $f_{D|X}(d_i|x_i; \theta)$.
2. Use estimated $\hat{\theta}$, with the assumed density function used for Step 1, to calculate the *observed* and *unobserved* GPS values: $\hat{R}_i = f_{D|X}(d_i|x_i; \hat{\theta})$ and $\hat{R}_{d,i} = f_{D|X}(d|x_i; \hat{\theta})$, for all d of interest.
3. Using $\hat{R}_{d,i}$ find a region of common support $\mathcal{C} \subseteq \mathcal{D}$, and check that balance holds for this region. We denote the number of observation in the common support region by S .
4. Estimate $\mathbb{E}(Y_i|D_i, \hat{R}_i)$ using a flexible regression model with parameters α : $m(y_i|d_i, \hat{R}_i; \alpha)$.
5. Average over predicted values from step 4., evaluated at dose d , to obtain a point estimate of the expected response at d : $\hat{\mu}(d) = S^{-1} \sum_{i=1}^S \mathbb{E}_{Y|D,R}[Y_i(d)|d, \hat{R}_{d,i}, \hat{\alpha}]$. Repeat for all doses of interest.
6. Use a single bootstrap resampling scheme over Steps 1 to 5 for variance estimation.

Details on each step of the algorithm are included in the online supplementary materials.

2.2 PS Construction

The success of PS estimators relies on having an observed covariate vector which is sufficient to represent confounding. Ignorability and the balancing property of the PS will not necessarily hold given only a subset of confounders, and consequently, the PS may not eliminate all sources of bias (e.g., Pearl 2000, 2009). In the context of binary treatments, Rubin (1997) argues against inclusion of nonconfounding covariates in the PS, even if they have an important influence on assignment, because it can have an adverse effect on efficiency of causal models. In this subsection, we consider this issue in relation to the GPS as it has important implications for the mixed model approach that we propose below.

We refer to a GPS that conditions on confounding and nonconfounding covariates as a fully conditional GPS (FCGPS). Let $r^*(d, x_i, u_i) = f_{D|X,U}(d|x_i, u_i)$ represent a value from the conditional density of $D_i = d$ given both X_i and U_i , which can be used to form the FCGPS random quantities $R_i^* = r^*(D_i, X_i, U_i)$ and $R_{d,i}^* = r^*(d, X_i, U_i)$ for fixed d . There are five properties of the FCGPS relevant for dose-response estimation that we state here. Detailed results and proofs for each of these are available in supplementary material.

- i. The FCGPS retains the balancing property such that: $I_{D_i}(d) \perp\!\!\!\perp X_i, U_i | r^*(d, X_i, U_i)$.
- ii. Ignorability can be established conditional on the FCGPS rather than covariate vector X_i : $Y_i(d) \perp\!\!\!\perp I_{D_i}(d) | r^*(d, X_i, U_i)$.

- iii. If assignment to the treatment is ignorable given baseline characteristics X_i , then conditioning on the FCGPS eliminates biases associated with differences in those confounding covariates. Consequently, the dose-response function can be computed

$$\begin{aligned}\mu(d) &= \mathbb{E}[Y_i(d)] = \mathbb{E}_X[\mathbb{E}[Y_i(d)|X_i]] \\ &= \mathbb{E}_{X,U}[\mathbb{E}[Y_i(d)|I_{D_i}(d), r^*(d, X_i, U_i)]] \\ &= \mathbb{E}_{X,U}[\mathbb{E}[Y_i|I_{D_i}(d), r^*(d, X_i, U_i)]].\end{aligned}$$

- iv. Estimates of the dose-response function based on the true (data generating) GPS will be *at least as efficient* as those based on the FCGPS.
- v. The FCGPS is drawn from a conditional density for d that has smaller variance than the conditional density from which the GPS is drawn. The implication of this result is that, at any level of the treatment, the FCGPS conditional density will have greater mass concentrated over a smaller range of units. Consequently, the common support region established using the GPS will be at least as large as that using the FCGPS.

In short, the FCGPS retains the properties of balancing and ignorability and will therefore still provide unbiased, although potentially less efficient, estimates of the dose-response function under correct model specification. However, redundant conditioning may reduce the size of the sample available to estimate causal quantities because it will be more difficult to find common support over the covariate distributions.

2.3 Mixed Models for PS Estimation in the Multi-Period Setting

So far we have assumed that the vector of observed covariates X_i is sufficient to represent confounding but in practice some confounding covariates may be unobserved, or even unknown. In these cases, and particularly when longitudinal data are available, we propose use of a mixed model approach to adjust for time-invariant unobserved confounding. We focus here on the case of a Gaussian treatment where a linear mixed model (LMM) is appropriate. Similar results for non-Gaussian treatments can be derived using the same principles for generalized linear mixed models (GLMMs).

A LMGPS model for a longitudinal data structure, comprising N units (or subjects), $i = 1, \dots, N$, each of which has n_i measurements made over times t , $t = 1, \dots, n_i$, giving a total of $n = \sum_{i=1}^N n_i$ sample observations, takes the form $D_{it} = X_{it}^T \theta_1 + Z_{it}^T \theta_{2i} + e_{it}$, where $X_{it}^T \theta_1$ is the fixed effects part of the model with p -dimensional design vector X_{it} and parameter vector θ_1 , Z_{it} is the q -dimensional design matrix for the vector of random effects (RE) $\theta_{2i} \sim \mathcal{N}(0, \mathbf{G})$, and $e_i^T = (e_{i1}, \dots, e_{in_i}) \sim \mathcal{N}(0, \mathbf{H}_i)$ is a vector of random errors. Using bold to denote matrices, \mathbf{G} and \mathbf{H}_i are $(q \times q)$ and $(n_i \times n_i)$ positive definite covariance matrices of RE θ_{2i} and error vector e_i for unit i respectively, and we assume that θ_{2i} and e_i are independent within subjects and θ_{2i} and e_{it} are independent across subjects given the covariates (see Diggle et al. 2002, for a discussion of these assumptions and the conditions under which they may be violated).

The RE component of LMMs can essentially be used as a target of inference in one of two ways: in relation to the population from which the RE is drawn, or in relation to the prediction of the realized value of the RE for unit i (for a discussion of this distinction see Ruppert, Wand, and Carroll 2003; McCulloch, Searle, and Neuhaus 2008; Fitzmaurice and Molenberghs 2009; Raudenbush 2009). The former model may be interpreted via marginalization as $D_i|\mathbf{X}_i \sim \mathcal{N}(\mathbf{X}_i\theta_1, \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i^\top + \mathbf{H}_i)$ and for the latter the interpretation is $D_i|\theta_{2i}, \mathbf{X}_i, \mathbf{Z}_i \sim \mathcal{N}(\mathbf{X}_i\theta_1 + \mathbf{Z}_i\theta_{2i}, \mathbf{H}_i)$, where $D_i^\top = (D_{i1}, \dots, D_{in_i})$ is an n_i dimensional vector of observed doses, \mathbf{X}_i is $(n_i \times p)$ and \mathbf{Z}_i is $(n_i \times q)$.

2.4 A Random Intercepts GPS Model to Address Time-Invariant Confounding

For GPS estimation in the presence of unobserved confounding, we propose a subject-specific interpretation of the mixed model, $D_{it} = X_{it}^\top\theta_1 + b_i + e_{it}$, where $b_i \sim \mathcal{N}(0, \sigma_b^2)$ is a random intercept. By specifying random unit level intercepts within the GPS, rather than the outcome model, they are by construction rendered correlated with treatment assignment and therefore potentially useful in representing sources of time-invariant unobserved confounding. The LMGPS approach can be further generalized in two ways. First, we can relax the implicit assumption that the RE are uncorrelated with the covariates (i.e., $\mathbb{E}[b_i|X_{it}] = 0$) by using a *mean centered* model, which adds time-averaged values of the covariates (i.e., $\bar{X}_{i,1} = n_i^{-1} \sum_{t=1}^{n_i} X_{it,1}$) to the model yielding individual correlated RE $b_i = \bar{X}_i^\top \xi + \omega_i$ with parameter vector ξ and random component $\omega_i \sim \mathcal{N}(0, \sigma_\omega^2)$, such that $\mathbb{E}[b_i|X_{it}] = \bar{x}_i^\top \xi$. Second, we can introduce *dynamics* to allow confounding from lagged doses by specifying a potentially autoregressive error term $\varepsilon_{it} = \rho\varepsilon_{it-1} + v_{it}$, with $|\rho| < 1$ and $v_{it} \sim \mathcal{N}(0, \sigma^2)$. This dynamic model, with either correlated or uncorrelated RE, is equivalent to an autoregressive distributed lag (ADL) model of order (1,1), that is

$$\begin{aligned} D_{it} &= X_{it}^\top\theta_1 + b_i + \varepsilon_{it} \\ &= \rho D_{it-1} + X_{it}^\top\theta_1 - \rho X_{it-1}^\top\theta_1 + b_i(1 - \rho) + v_{it}, \end{aligned}$$

from which we can recover an estimate of dynamic confounding (ρ). If appropriate the response history through time $t - 1$, which we denote $\mathcal{H}_i^Y(t - 1)$, can be added as an exogenous predetermined covariate in the LMGPS models.

Specification of RE in the GPS model can help capture time-invariant unobserved confounding and therefore improve the plausibility of the ignorability assumption. Using the GPS algorithm, with a mixed model specification for Step 1, we can estimate two key parameters of interest: the APO, $\mu(d) = \mathbb{E}_{it}[Y_{it}(d)]$, and the ATE, $\tau(d) = \mathbb{E}_{it}[Y_{it}(d)] - \mathbb{E}_{it}[Y_{it}(0)]$. However, since the predicted RE in the LMM cannot distinguish between sources of variation that arise from time-invariant confounding or nonconfounding characteristics, adverse effects on efficiency and evaluation of common support may arise as discussed in Section 2.2. It is also worth noting that in adopting an LMM approach we are obliged to assume a model for the RE, which if misspecified can adversely effect estimation of random and fixed parameters. In applied work, for instance, normality of RE is typically assumed for computational convenience

but in some applications could result in a misspecified model. A summary of the consequences of misspecification and some potential solutions via nonparametric methods were reviewed by Huang (2011) while Abad, Litire, and Molenberghs (2010) proposed some useful diagnostics tests.

2.5 Comparable Approaches

There are other prominent approaches for estimation of continuous treatment effects in the longitudinal setting that have been discussed in the literature and that could potentially address time-invariant unobserved confounding.

1. Outcome regression (OR)—standard longitudinal OR approaches, using correlated RE or fixed effects, can be used to address time-invariant confounding (see Diggle et al. 2002). If the OR model is correctly specified and there is overlap in the sample such approaches will tend to be more efficient than the multistep GPS approach. Key advantages of the GPS over OR models arise in working with a scalar value rather than a potentially high-dimensional covariate vector. This allows for effective use of flexible approaches in modeling potentially nonlinear dose-response functions, for instance via semiparametric and polynomial regressions with interaction terms. Moreover, it is highly effective in isolating the region of common support, a task that is difficult using multiple covariates (for discussion see Joffe and Rosenbaum 1999). The GPS approach is of course more involved, with multiple modeling steps, but these generate additional statistical summary information which can be highly informative both for the inference procedure and for substantive or policy issues.
2. Difference-in-differences (DID) GPS (DIDGPS)—Flores et al. (2012) implemented a difference-in-differences GPS estimator which transforms the response into before-after outcomes and then applies the GPS approach for confounding adjustment. This is a continuous treatment counterpart to the conditional generalized DID estimator for binary treatments proposed by Heckman, Ichimura, and Todd (1997) and Heckman et al. (1998). Differencing outcomes allows for time-invariant unobservable factors to influence selection and consequently reduces bias when the data are contaminated by temporally invariant components. Since the RE are differenced out rather than explicitly specified, DID approaches do not require an assumed model for RE as in the LMM approach. Unlike the LMGPS, however, this estimator requires that the time-invariant unobserved confounders be related to the response linearly and that there be pretreatment observations available. Furthermore, the extension to longitudinal treatments is not straightforward, and also the approach sacrifices some efficiency via data loss through differencing.
3. Multivariate GPS (MGPS)—Moodie and Stephens (2012) proposed a repeated measures, or multivariate, longitudinal generalization of the GPS approach which analyses the effect of treatments applied over time allowing for confounding from lagged doses and responses. They showed that sequential conditional independence (i.e., by-period conditional independence of treatment and response given

response and treatment histories) can be established using the MGPS. The MGPS approach is not explicitly designed to address time-invariant confounding but since the lagged doses and responses are in turn influenced by time-invariant RE it could reduce bias from these sources.

4. Marginal structural models (MSMs)—MSMs (Hernán, Brumback, and Robins 2001) have been used extensively to analyze treatment effects in the longitudinal setting. The approach is explicitly formulated to address the problem of time varying confounding using the underlying principle of inverse probability weighting to create pseudo-populations. Primarily, MSMs have been used to estimate the *total* effect of a treatment regime on an end-of-study outcome, but they can also be used to estimate the *direct* treatment effect as an average value over all periods (for a discussion see VanderWeele 2009). Weighting-based approaches such as MSMs, are quite different from the regression-based approaches that we adopt. As we have shown, the mixed model approach can help to account for unmeasured confounding in regression, yet to incorporate such structures into weighting-based approaches requires further methodological investigation.

In the next section, we present simulations to illustrate the properties of the LMGPS approach in relation to other comparable longitudinal estimators for continuous treatments.

3. SIMULATIONS

Our simulations are conducted on samples of 1000 observations comprising 100 subjects each observed at 10 time points. We index subjects by i , $i = (1, \dots, N)$ and time points by t , $t = (0, \dots, n_i)$ giving a total of $n = \sum_{i=1}^N n_i$ sample observations. There are three normally distributed independent covariates: $X_{1it} \sim \mathcal{N}(0.2, 0.1)$, $X_{2i} \sim \mathcal{N}(1.0, 0.6)$, and $U_i \sim \mathcal{N}(0.2, 0.1)$. Covariate X_{1it} varies over time, but is not a function of time, while X_{2i} and U_i are time-invariant.

We specify the following relationships between the covariates and the treatment assignment D and response Y :

$$Y_{it}(d)|X_{1it}, X_{2i} \sim \mathcal{N}(20 \exp(d[X_{1it} - 0.25X_{2i}]), \sigma_Y^2)$$

$$D_{it}|X_{1it}, X_{2i}, U_i \sim \mathcal{N}(1.0 + 4.0X_{1it} + 2.0X_{2i} + U_i, \sigma_D^2).$$

Thus, X_1 and X_2 are confounders for D in a nonlinear relationship with Y , while U is a nonconfounding covariate. Note that due to the nonlinear nature of the relationship between Y and X_2 , differencing will not eliminate time-invariant confounding.

The simulation set up is chosen to demonstrate the dual challenge of addressing time-invariant unobserved confounding with a nonlinear dose-response relationship. We assume that the analyst is ignorant of the data-generating process. The following models for the dose-response are tested

1. $\hat{\mu}(d)_{\text{OR1}}$ —a linear OR model based on an incorrectly specified generalized linear model (GLM), with erroneous exclusion of the time-invariant confounder X_{2i} : $\mathbb{E}[\log(Y_{it})|D_{it}, X_{1it}] = \beta_0 + \beta_1 D_{it} + \beta_2 X_{1it}$. The dose-response is derived by taking the mean of the predicted values of this model evaluated at $D_{it} = d$.

2. $\hat{\mu}(d)_{\text{OR2}}$ —as model OR1 but with RE (η_i) specified for each subject: $\mathbb{E}[\log(Y_{it})|D_{it}, X_{1it}, \eta_i] = \beta_0 + \beta_1 D_{it} + \beta_2 X_{1it} + \eta_i$.
3. $\hat{\mu}(d)_{\text{OR3}}$ —as model OR2, but estimated using a flexible generalized additive mixed model (GAMM) specification
4. $\hat{\mu}(d)_{\text{GPS}_t}$ —a GPS dose-response estimator based on a correctly specified linear GPS model: $\mathbb{E}[D_{it}|X_{1it}, X_{2i}] = \theta_0 + \theta_1 X_{1it} + \theta_2 X_{2i}$.
5. $\hat{\mu}(d)_{\text{GPS}_f}$ —a GPS dose-response estimator based on an incorrectly specified GPS model, with erroneous exclusion of the time-invariant confounder X_{2i} .
6. $\hat{\mu}(d)_{\text{FCGPS}}$ —a GPS dose-response estimator based on an incorrectly specified GPS model, with inclusion of covariates X_{1it} , X_{2i} , and U_i .
7. $\hat{\mu}(d)_{\text{LMGPS}}$ —a linear mixed GPS model dose-response estimator, with exclusion of covariate X_{2i} , but inclusion of unit level RE.
8. $\hat{\mu}(d)_{\text{MGPS}}$ —multivariate GPS Markov model which excludes time-invariant confounder X_{2i} , but adjusts for lagged values of dose and response: $\mathbb{E}[D_{it}|X_{1it}, Y_{it-1}, D_{it-1}] = \theta_0 + \theta_1 X_{1it} + \theta_2 Y_{it-1} + \theta_3 D_{it-1}$.

The parameters being estimated are population APOs. We impose common support in the GPS based models to illustrate the effects of conditioning on sample size, but in our particular set up it does not affect the value of the estimated parameters. For GPS models, the conditional density of the treatment given the covariates is estimated using a Gaussian GLM or GLMM, observed and unobserved GPSs are calculated via the normal density, and the conditional expectation of the log of the outcome given D and \hat{R} is approximated using a GAM approximation, with smooth main effect functions for D and R and a smooth interaction. Using a method detailed in the supplementary materials we selected the largest common support region in which the estimated probability of assignment to each dose ≥ 0.01 . All models are specified with a log transformation of the outcome and results are presented on the scale of the (log) response. The simulations are run on 1000 generated datasets of size 1,000. The mean dose is 4.0 and APO estimates are calculated for doses 3, 4, and 5. Table 1 reports average estimates (Av Est), average estimated variances (Av Est Var), empirical variances (Emp Var) calculated via 1000 bootstrap replications, mean squared error (MSE), coverage based on bootstrap bias-corrected adjusted 95% confidence intervals, and the size of common support regions (S).

The first column of numbers in Table 1 confirms that the size of the common support region will tend to reduce with more extensive conditioning (i.e., FCGPS and LMGPS) relative to the true GPS (GPS_t), and is underestimated by GPS_f and MGPS. Note that average estimated variances and empirical variances are much the same for models OR1, OR2, OR3, and GPS_f; but not for the remaining models. This is because the imposition of common support reduces sample size and consequently inflates estimated variances, but it does not affect bias in this particular simulation set up and so the empirical variance of estimates across datasets is unaffected.

The OR dose-response estimates are biased and correspondingly have poor coverage. There are two effects at play here. For all specifications (OR1, OR2, and OR3) the effect of the

Table 1. Simulation results for a Gaussian nonlinear dose* response function: average estimate, average estimated variance, empirical variance, MSE, and coverage (nominal level is 95%)

		S	dose = 3.00	dose = 4.00	dose = 5.00
truth			2.846	2.796	2.746
OR1	Av Est		3.290	2.903	2.517
	Av Est Var		0.006	0.007	0.013
	Emp Var	1000	0.007	0.008	0.014
	MSE		0.205	0.019	0.067
	Coverage		0.3	76.8	41.7
OR2	Av Est		3.159	2.903	2.647
	Av Est Var		0.008	0.009	0.014
	Emp Var	1000	0.007	0.008	0.014
	MSE		0.106	0.020	0.024
	Coverage		1.8	81.4	67.0
OR3	Av Est		3.016	2.779	2.593
	Av Est Var		0.008	0.009	0.014
	Emp Var	1000	0.009	0.011	0.017
	MSE		0.038	0.011	0.041
	Coverage		23.9	92.3	58.3
GPS _t	Av Est		2.814	2.802	2.795
	Av Est Var		0.021	0.026	0.050
	Emp Var	638	0.016	0.022	0.039
	MSE		0.017	0.022	0.041
	Coverage		94.2	96.2	94.6
GPS _f	Av Est		2.926	2.767	2.497
	Av Est Var		0.015	0.023	0.023
	Emp Var	994	0.014	0.024	0.024
	MSE		0.021	0.025	0.085
	Coverage		90.3	92.3	63.2
FCGPS	Av Est		2.814	2.802	2.793
	Av Est Var		0.023	0.026	0.053
	Emp Var	602	0.017	0.022	0.042
	MSE		0.018	0.022	0.044
	Coverage		95.8	95.7	94.7
LMGPS	Av Est		2.823	2.794	2.742
	Av Est Var		0.025	0.027	0.056
	Emp Var	608	0.017	0.022	0.045
	MSE		0.017	0.022	0.045
	Coverage		96.2	95.7	94.6
MGPS	Av Est		2.853	2.787	2.676
	Av Est Var		0.019	0.024	0.040
	Emp Var	743	0.015	0.022	0.037
	MSE		0.015	0.022	0.041
	Coverage		95.8	95.4	92.6

treatment is confounded due to the erroneous omission of the time-invariant confounder X_2 . The inclusion of RE in OR2 and OR3 fails to adjust for this source of confounding, since by construction, the RE are independent of the covariates and therefore capture time-invariant unobserved variables that are strictly non-confounding. The second effect, relevant to OR1 and OR2, is that the semi-log-linear model specification fails to provide a good approximation to the true nonlinear nature of the dose-response function. With a high-dimensional covariate vector this effect may be evident due to the limited potential for use of highly flexible functional forms.

With the exception of model specification GPS_f, the GPS models generally perform well in estimating the dose-response and coverage is good, even though these model still involve

approximations to the true nonlinear dose-response function. This is because they offer a high degree of flexibility in approximating the unknown function while providing adequate adjustment for confounding covariates. Note that use of a best linear unbiased predictor (BLUP) to calculate the LMGPS gives similar results to those obtained for the FCGPS, even though this model does not include the time-invariant confounder explicitly. This is because the RE in LMGPS are correctly located in the sense that they do adjust for confounding, in addition to other sources of unit level variance. Through inclusion of lags in both treatment and response the MGPS is reasonably robust to the exclusion of time-invariant RE, but this estimator is not designed to address this particular issue directly and consequently we find that it does not perform as well as the LMGPS far from the mean dose. Excluding time-invariant confounding covariates from GPS estimation, as in the GPS_f specification, produces poor estimates of the dose-response.

4. CASE STUDY: QUANTIFYING CAUSAL EFFECTS OF ROAD NETWORK CAPACITY EXPANSIONS ON TRAFFIC VOLUME AND DENSITY

We now apply the method outlined above to our longitudinal case study on the effect of road network capacity expansion on aggregate traffic volumes. The data available for analysis describe mobility and traffic conditions for 101 major U.S. cities over the period 1985 to 2010. They are collated by the Texas Transportation Institute (TTI) at the University of Austin Texas and are publicly available with full documentation at <http://mobility.tamu.edu/ums/>. A description of the data and a table of summary statistics is provided in the supplementary materials.

The available longitudinal data allow us to represent three key features of our problem: response (y_{it}), treatment (d_{it}), and confounding covariates (x_{it}). We define our response variable as annual proportional change in aggregate urban traffic volume, with volume measured by annual vehicle miles traveled (vmt) in each city i : $y_{it} = vmt_{it}/vmt_{it-1}$. Our “treatment” variable is a measure of the proportional change in urban road lane miles (lms) in each year: $d_{it} = lms_{it}/lms_{it-1}$. The advantage of using proportional changes for treatment and response is that, since we can condition on initial scale in the GPS regression, we can obtain APO estimates that have an intuitive and general interpretation as the proportional change in vmt caused by a proportional increase in capacity.

The previous literature offers guidance on likely sources of confounding. To model time-varying confounders we use the available longitudinal data to construct the following covariates, each measured in the year immediately prior to the treatment: (i) lagged-response (to allow for reverse causality), (ii) congestion (*annual hours of delay per vmt*), (iii) traffic volume (vmt), (iv) network scale (lms), (v) network composition (*freeway lms / arterial lms*), (vi) traffic composition (*arterial vmt / freeway vmt*), (vii) mode share (*annual passenger miles traveled by public transport*), (viii) productivity (*metropolitan wage rate in dollars per annum*), (ix) economic structure (*metropolitan share of manufacturing jobs*), (x) employment and population distribution and growth (covariates measuring the levels and proportional growth), (xi) personal income (*average metropolitan personal income in dollars per annum*), and (xii) state fuel

price (*dollars per gallon*). Further detail on the hypotheses underpinning inclusion of these covariates, and summary statistics for the data, can be found in the supplementary materials.

As discussed in the introduction, the previous literature tends to use model specifications that allow for confounding from unobserved time-invariant city level characteristics. These could include physical layout, topography, climate, geography, cultural factors, and features of network engineering. Thus, in addition to the time-varying confounders we draw on a mixed model framework for the GPS. We use a mean centered LMGPS which, as explained in Section 2, provides the most general specification in that it allows for correlation between the time-invariant RE and the time varying covariates. By virtue of the fact that a lagged response features as a covariate in the GPS model, correlation with the RE should be assumed. Our mean centered LMGPS thus accommodates the key issues surrounding induced demand estimation raised in the literature. For comparison and testing, a number of other GPS based models and conventional longitudinal models are estimated. We summarize these results below and provide full tables and statistics in the appendix.

The LMGPS approach was illustrated through simulation in the previous section. In the application an identical procedure is followed. We first estimate the conditional density of treatment given covariates. To allow flexibility of form, we use a Gaussian penalized spline model with smooth main effect functions for each covariate and automatic knot selection, estimated using the GAMM4 package in R. Estimation is by restricted estimation maximum likelihood (REML) and approximate results on significance are given in Table 2. The p -values, which derive from a Wald test of individual smooth terms for equality to zero, are based on a statistic with an approximate F distribution. As described by Wood (2006) the p -values should be interpreted with caution as they are typically lower than they should be when the null is true creating a tendency toward over-rejection of the null. Using the Bayesian information criterion (BIC) of this model as a guide we found that a more parsimonious specification could be achieved by dropping some covariates related to population and employment distribution and growth. The final set of covariates used is shown in Table 2.

Using the method detailed in the appendix we select a region of common support in which the estimated probability of assignment to dose ≥ 0.01 . This yields a region of 1353 comparable observations. The large exclusion of observations ($\approx 45\%$) required to achieve comparability implies that estimated APOs and ATEs will likely differ from the corresponding population parameters. Tables A1 and A2 in the appendix provide detail on differences in characteristics between the common support region and the full sample. Most noticeable is that smaller cities are somewhat under-represented in the common support region and consequently we find that the mean values for traffic volume, network scale, productivity, population, and employment are larger than for the full sample.

We next test for the balancing property of the GPS. In a similar manner to Hirano and Imbens (2004) and Flores et al. (2012), we regress the treatment on the covariates, the GPS, and a set of indicator variables corresponding to the following treatment discretization: [1,1.005], (1.005,1.01], (1.01,1.015], (1.015,1.02], (1.02,1.025], (1.025,1.03], (1.03,1.035], (1.035,1.04], (1.04,1.045], (1.045,1.05], (1.05,1.1]. The

Table 2. REML Gaussian penalized spline model results for the conditional density of the treatment given covariates. Columns contain test statistics and p -values for terms in the spline model

	F	p -value
Lag response	10.844	2.56E-12
Congestion	11.488	0.000
Traffic volume	2.895	3.00E-02
Network scale	10.775	2.61E-05
Network composition	2.315	0.12823
Traffic composition	4.599	6.11E-05
Economic structure	1.775	0.183
Productivity	2.855	0.034
Mode share	1.850	0.121
Employment	3.182	0.008
Fuel price	3.398	0.020
Personal income	2.606	0.011
Population growth	1.924	0.082
Year	7.604	1.65E-08
Deviance	1.28E-04	
BIC	-14245	
n	2474	

BIC values obtained from the linear regression models with and without covariates, -21371 and -21414 , respectively, indicate that the inclusion of covariates leads to a deterioration in model adequacy. We also conduct an F -test between the restricted (without covariates) and unrestricted models giving an F statistic (p -value) of 1.250 (0.264). These results suggest that the balancing property has been achieved for our GPS specification.

Next, we approximate the conditional mean of the outcome given treatment level and the estimated GPS. We do so using a Gaussian penalized spline GAM estimated by REML, with automatic knot selection and smooth main effect functions for d and \hat{R} . To test the validity of this simple specification, rather than use a rectangular grid of knot points in the (d, \hat{R}) plane, we choose knots for the \hat{R} component tailored to a range discretization of d into 10 strata. The BIC of this model (-6416) does not support the need for separate smooths, and in fact the model generates very similar dose-response estimates to those obtained with a single smooth. This test provides only a rudimentary check for model specification, but it does also demonstrate the considerable flexibility in form offered via the GPS approach relative to OR approaches with many covariates.

We do not observe the set of potential outcomes across doses of interest. Instead, we construct the dose-response function using the conditional mean-response model by taking the mean of its predicted values on the response scale, evaluated at $D = d$ and $\hat{R}_{d, it} = r(d, x_{it})$. We calculate the dose-response for increments of 0.0025 over the range of doses from 1.0 to 1.05 (i.e., from 0% to 5% capacity expansion). Results for selected doses are shown in Table 3 with bootstrap standard errors. The bootstrap calculations are based on 500 block replications given both the GPS estimation and the averaging required to estimate the dose-response. In addition to the dose-response estimates, we also report ATEs defined as the expected proportional change in response under any given capacity expansion net of the expected proportional change that would occur under no capacity

Table 3. Estimates of the dose-response function and associated ATEs with bootstrapped standard errors

Dose	$\hat{\mu}(d)$	S.E.	$\hat{\tau}(d)$	S.E.	Elasticity	S.E.
1.000	1.014	(0.003)	—	—	—	—
1.005	1.020	(0.001)	0.007	(0.003)	1.393	(0.539)
1.010	1.026	(0.001)	0.012	(0.003)	1.217	(0.295)
1.015	1.028	(0.001)	0.015	(0.003)	0.997	(0.206)
1.020	1.030	(0.002)	0.017	(0.003)	0.827	(0.147)
1.025	1.032	(0.002)	0.018	(0.003)	0.720	(0.125)
1.030	1.033	(0.003)	0.019	(0.004)	0.645	(0.129)
1.035	1.034	(0.003)	0.020	(0.004)	0.577	(0.117)
1.040	1.034	(0.004)	0.021	(0.004)	0.518	(0.112)
1.045	1.035	(0.005)	0.021	(0.005)	0.472	(0.117)
1.050	1.035	(0.005)	0.022	(0.006)	0.435	(0.112)

expansion. We estimate ATEs using

$$\hat{\tau}(d) = \frac{1}{n} \sum_{i=1}^N \sum_{t=1}^{n_i} [\hat{Y}_{it}(d)] - \frac{1}{n} \sum_{i=1}^N \sum_{t=1}^{n_i} [\hat{Y}_{it}(1)],$$

which with $y_{it} = vmt_{it}/vmt_{it-1}$ is a discrete approximation to $E[\partial \log vmt_{it}(d)] - E[\partial \log vmt_{it}(1)]$.

Table 3 shows our estimated responses and ATEs at a selection of doses. The response value corresponding to dose 1.000 represents average “natural growth” in traffic (i.e., average growth under no treatment), which we estimate to be 1.4% per annum. All other average response estimates shown in the table, which are associated with network capacity increases of various doses, indicate evidence of induced road traffic demand having adjusted for observed time-varying and unobserved time-invariant confounders and having ensured common support. The bootstrapped standard errors indicate statistically significant effects pointwise at all doses.

Figure 1 illustrates the ATEs graphically with bootstrapped 95% pointwise confidence intervals. Note that the ATEs are increasing with dose across the range of treatment considered, and the effect is statistically significant at all doses. This implies that city dwellers have tended to experience larger traffic volumes due to network expansions with potentially adverse implications for pollution and the incidence of collisions.

We can transform our ATEs to the *elasticity* scale typically reported in the existing literature by dividing the ATE estimate by the corresponding proportional change in lane miles. For our LMGPS model, this gives a range from 1.393 (0.539) to 0.435 (0.112) with a mean of 0.772 (0.170). The elasticities decrease from the smallest to largest dose. Existing empirical studies quote single point elasticity estimates, typically under 0.700, although the recent study by Duranton and Turner (2011) which uses IV as a causal method reports a larger elasticity of around 1.000. Our results indicates that there may be considerable heterogeneity in response across different doses, and therefore that there is value in adopting a causal dose-response approach rather than a conventional OR approach for a single point estimate.

The results can also be used to indicate the effect of capacity expansions on traffic densities by taking the ratio of response to doses. Our estimates suggest that, given the combined effects of natural growth and induced demand, increases in network capacity of less than 3.5% may not cause traffic densities to reduce. Furthermore, for our original sample of 101 cities over

26 years, only 533 observation (20%) have experienced annual growth in lane miles greater than 3.5%. So, the induced demand effect suggested by our estimates really is substantial, implying that even major network capacity expansions may not actually cause traffic densities to fall, and by extension, may do little to ameliorate urban congestion.

For illustrative purposes and to provide further insight into the LMGPS methodology, we generate additional results from alternative model specifications proposed in the literature. These results are summarized below and presented in full in the appendix.

Alternative GPS model specifications. We estimate a GPS model with covariates but without adjustment for time-invariant confounding, and an MGPS model as described above. These models, which do not adjust directly for time-invariant confounding, give substantially similar results to the LMGPS estimates, but the estimated induced demand effects is larger. The GPS model gives a mean elasticity of 0.804 (0.163) and the MGPS model 1.128 (0.166).

Models using conventional panel OR approaches. We estimate the following models: pooled OLS (POLs) with and without covariates, RE, fixed effects (FE), first difference (FD), and dynamic-panel generalized method of moments (DPGMM) (see Hall 2005). In all cases, the treatment is specified in quadratic form. If we assume that the covariate vector correctly represents time-varying confounding, then the POLs model without covariates will give inconsistent estimates, the POLs with covariates and the RE models will give consistent estimates if there is no time-invariant confounding, and the FE and FD models will give consistent estimates in the presence of both time-varying and time-invariant confounding, as will the DPGMM model but it also allows for confounding from lagged values of the response. Note that common support is not imposed in these conventional approaches implying two important points for interpretation and comparison: first, the causal parameter being estimated may differ from the corresponding value for the common support region; second, parameter estimates are calculated without concern for the comparability of units and thus a key requirement of the potential outcome framework for causal inference is not met.

The POLs model without covariates makes no adjustment for confounding and gives a mean elasticity of 1.121 (0.064), somewhat higher than values typically reported in the literature. The POLs model with covariates and the RE model give mean elasticities that are broadly similar to the mean value for the LMGPS model, 0.769 (0.096) and 0.785 (0.086) respectively, but in fact the pattern of ATE estimates from these models is substantially different indicating considerably less heterogeneity in treatment effect by dose. The FE, FD, and DPGMM models, which adjust for time-invariant confounding, produce results that are similar in magnitude to those reported in the literature with mean elasticities of 0.530 (0.084), 0.597 (0.090), and 0.611 (0.296) respectively, but again show less heterogeneity in ATEs than indicated by the GPS based models.

All estimates from the additional models indicate evidence of induced road traffic demand from network expansion, but the magnitude of estimated effects varies. Adjustment for time varying confounding appears to produce lower estimates and

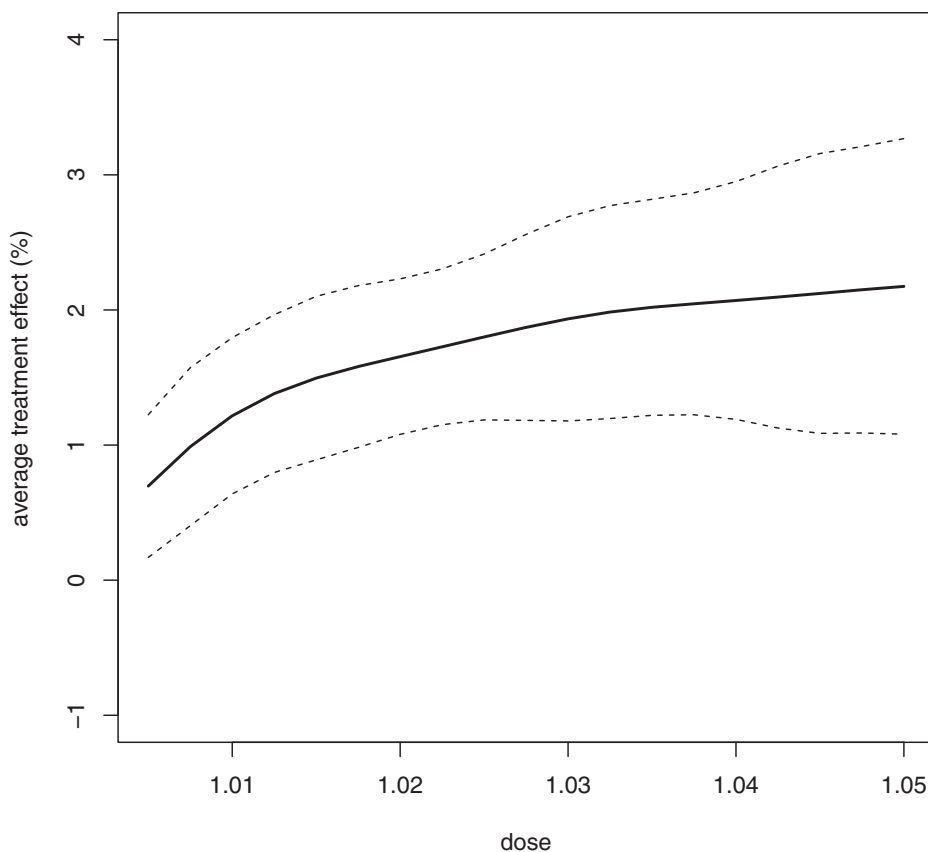


Figure 1. Average treatment effects (%) by “dose”—network capacity expansion—in the traffic volume example.

additional adjustment for time-invariant confounding appears to reduce the estimates still further. The FD and FE panel data models show improved explanatory power, as measured by adjusted R^2 , and the Hausman test selects the FE over RE model. These diagnostics are consistent with the existence of time-invariant confounding. However, while the panel models provide a useful perspective which is broadly supportive of the LMGPS approach we propose, since they do not address common support and are based on a restrictive a priori assumption about the form of the dose-response, direct comparison of results is inhibited.

In summary, under the assumptions of our model, we find evidence of a positive causal relationship between network capacity expansions and traffic volumes. The scale of the effect indicates that network capacity expansions may do little to reduce traffic densities other than in extreme cases (i.e., expansions of 3.5% and above). This implies that as a remedy for congestion, capacity expansion is at best a risky strategy and city authorities should be aware of the potential scale of the induced demand effect.

5. CONCLUSIONS

In this article, we have studied the effect of road network capacity expansions on aggregate traffic volumes and densities in U.S. cities using a linear mixed model GPS approach for continuous dose-response estimation. Given the assumptions of our model, our results suggest that capacity expansions can give rise to a direct increase in aggregate urban traffic volumes, and we find that the effect may be substantial such that even major

capacity increases can actually lead to little or no reduction in network traffic density. One implication is that by building more roads in major urban areas we may create increasing pollution, congestion, collisions, and other negative consequences of urban motoring. On the other hand, we may also permit more mobility for urban dwellers.

GPS based estimators are attractive because they permit the use of flexible approaches that make minimal a priori assumption on the form of the dose-response relationship and they can be used to find a sample comprising observations that are broadly comparable for treatment effect estimation. The methodological key insight in this article is that by specifying RE within a GPS rather than mean-response model, they are rendered correlated with treatment assignment and therefore potentially useful in representing sources of unobserved time-invariant confounding. Our proposed estimator provides a general specification that could be useful in other areas of applied statistics as it accommodates three prevalent features of longitudinal problems: confounding from measured time-varying and unobserved time-invariant sources and bi-directionality between treatment and response.

Some limitations of our approach should be noted. First, since the predicted RE cannot distinguish between time-invariant unobserved heterogeneity that arise from confounding or non-confounding characteristics, their use in the PS model can potentially lead to more extensive conditioning than is strictly necessary for causal comparison. We have shown that while the inclusion of non-confounding characteristics in the PS model does not induce bias in the estimation of the dose-response, it

does give potentially less efficient estimates and can render the task of finding overlap in the covariate distribution more challenging. Second, in contrast to IV methods, we require “ignorability” to hold given the covariates, and while our approach will help address unmeasured *time-invariant* confounding, it will not eliminate bias from unmeasured time-varying confounding. Finally, two limitations that could be addressed in future research relate to the need for appropriate test statistics for model fit that are applicable for multistep estimation procedures, and the issue of measurement error and the propagation of error through the model components.

SUPPLEMENTARY MATERIALS

The supplementary material provides detail on the algorithm used for dose-response estimation, proofs for the properties of the mixed model GPS, a description of the data used in our application, and some additional results.

[Received April 2013. Revised July 2014.]

REFERENCES

- Abad, A. A., Litire, S., and Molenberghs, G. (2010), “Testing for Misspecification in Generalized Linear Mixed Models,” *Biostatistics*, 11, 771–786. [1443]
- Cervero, R. (2002), “Induced Travel Demand: Research Design, Empirical Evidence, and Normative Policies,” *Journal of Planning Literature*, 17, 3–20. [1440]
- (2003), “Road Expansion, Urban Growth, and Induced Travel: A Path Analysis,” *Journal of the American Planning Association*, 69, 145–163. [1440]
- Cervero, R., and Hansen, M. (2002), “Induced Travel Demand and Induced Road Investment: A Simultaneous Equation Analysis,” *Journal of Transport Economics and Policy*, 36, 469–490. [1440]
- Crump, R. K., Hotz, V. J., Imbens, G. W., and Mitnik, O. A. (2009), “Dealing With Limited Overlap in Estimation of Average Treatment Effects,” *Biometrika*, 96, 187–199. [1441]
- Diggle, P. J., Heagerty, P., Liang, K.-Y., and Zeger, S. L. (2002), *Analysis of Longitudinal Data*, Oxford: Oxford University Press. [1442,1443]
- Duranton, G., and Turner, M. A. (2011), “The Fundamental Law of Road Congestion: Evidence From U.S. Cities,” *American Economic Review*, 101, 2616–52. [1440,1447]
- Fitzmaurice, G., and Molenberghs, G. (2009), “Introduction and Historical Overview,” in *Longitudinal Data Analysis*, eds. G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs, London: Taylor & Francis, pp. 3–27. [1443]
- Flores, C. A., Flores-Lagunes, A., Gonzalez, A., and Neumann, T. C. (2012), “Estimating the Effects of Length of Exposure to Instruction in a Training Program: The Case of Job Corps,” *The Review of Economics and Statistics*, 94, 153–171. [1443,1446]
- Fulton, L. M., Meszler, D., Noland, R. B., and Thomas, J. (2000), “A Statistical Analysis of Induced Travel Effects in the U.S. Mid Atlantic Region,” *Journal of Transportation and Statistics*, 3, 1–14. [1440]
- Granger, C. W. J. (1969), “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods,” *Econometrica*, 37, 424. [1440]
- Hall, A. R. (2005), *Generalized Method of Moments, Advanced Texts in Econometrics*, Oxford: Oxford University Press. [1447]
- Hansen, M., and Huang, Y. (1997), “Road Supply and Traffic in California Urban Areas,” *Transportation Research Part A: Policy and Practice*, 31, 205–218. [1440]
- Heckman, J., Ichimura, H., Smith, J., and Todd, P. (1998), “Characterizing Selection Bias Using Experimental Data,” *Econometrica*, 66, 1017–1098. [1443]
- Heckman, J. J., Ichimura, H., and Todd, P. E. (1997), “Matching as an Econometric Evaluation Estimator: Evidence From Evaluating a Job Training Programme,” *Review of Economic Studies*, 64, 605–54. [1443]
- Hernán, M. A., Brumback, B., and Robins, J. M. (2001), “Marginal Structural Models to Estimate the Joint Causal Effect of Nonrandomized Treatments,” *Journal of the American Statistical Association*, 96, 440–448. [1444]
- Hirano, K., and Imbens, G. W. (2004), “The Propensity Score With Continuous Treatments,” in *Applied Bayesian Modeling and Causal Inference From Incomplete Data Perspectives*, eds. A. Gelman, and X. L. Meng, New York: Wiley, pp. 73–84. [1441,1446]
- Huang, X. (2011), “Detecting Random-Effects Model Misspecification via Coarsened Data,” *Computational Statistics & Data Analysis*, 55, 703–714. [1443]
- Hymel, K. M., Small, K. A., and Dender, K. V. (2010), “Induced Demand and Rebound Effects in Road Transport,” *Transportation Research Part B: Methodological*, 44, 1220–1241. [1440]
- Imbens, G. W. (2000), “The Role of the Propensity Score in Estimating Dose-Response Functions,” *Biometrika*, 87, 706–710. [1441]
- Joffe, M. M., and Rosenbaum, P. R. (1999), “Propensity Scores,” *American Journal of Epidemiology*, 150, 327–333. [1443]
- Kelly, F. (2008), “The Mathematics of Traffic in Networks,” in *The Princeton Companion to Mathematics*, eds. T. Gowers, J. Barrow-Green, and I. Leader, Princeton: Princeton University Press, pp. 862–870. [1440]
- McCulloch, C. E., Searle, S. R., and Neuhaus, J. (2008), *Generalized, Linear, and Mixed Models* (2nd ed.), New York: Wiley. [1443]
- Melo, P., Graham, D. J., and Canavan, S. (2013), “Effects of Road Investments on Productivity and Induced Travel Demand: Evidence for U.S. Urbanized Areas,” *Transportation Research Record*, 2297, 163–171. [1440]
- Moodie, E. M., and Stephens, D. A. (2012), “Estimation of Dose-Response Functions for Longitudinal Data Using the Generalized Propensity Score,” *Statistical Methods in Medical Research*, 21, 149–166. [1443]
- Noland, R. B. (2001), “Relationships Between Highway Capacity and Induced Vehicle Travel,” *Transportation Research Part A*, 35, 47–72. [1440]
- Noland, R. B., and Cowart, W. A. (2000), “Analysis of Metropolitan Highway Capacity and the Growth in Vehicle Miles of Travel,” *Transportation*, 27, 363–390. [1440]
- Pearl, J. (2000), *Causality — Models, Reasoning and Inference*, Cambridge, MA: Cambridge University Press. [1442]
- (2009), “Remarks on the Method of Propensity Score,” *Statistics in Medicine*, 28, 1415–1424. [1442]
- Raudenbush, S. W. (2009), “Targets of Inference in Hierarchical Models for Longitudinal Data,” in *Longitudinal Data Analysis*, eds. G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs, London: Taylor & Francis, pp. 167–186. [1443]
- Rubin, D. B. (1978), “Bayesian Inference for Causal Effects: The Role of Randomization,” *The Annals of Statistics*, 6, 34–58. [1441]
- (1997), “Estimating Causal Effects From Large Data Sets Using Propensity Scores,” *Annals of Internal Medicine*, 127, 757–763. [1442]
- Ruppert, D., Wand, M. P., and Carroll, R. J. (2003), *Semiparametric Regression*, Cambridge, MA: Cambridge University Press. [1443]
- Small, K., and Verhoef, E. (2007), *The Economics of Urban Transportation*, New York: Routledge. [1440]
- VanderWeele, T. J. (2009), “Marginal Structural Models for the Estimation of Direct and Indirect Effects,” *Epidemiology*, 20, 18–26. [1444]
- Wood, S. N. (2006), *Generalized Additive Models: An Introduction With R*, Boca Raton, FL: Chapman & Hall/CRC Press. [1446]