A Score-Driven Filter for Causal Regression Models with Time-Varying Parameters and Endogenous Regressors*

Francisco Blasques, Noah Stegehuis

Vrije Universiteit Amsterdam and Tinbergen Institute

Job Market Paper [link to latest version]

November 4, 2024

Abstract

This paper proposes a score-driven model for filtering time-varying causal parameters through the use of instrumental variables. In the presence of suitable instruments, we show that we can uncover dynamic causal relations between variables, even in the presence of regressor endogeneity which may arise as a result of simultaneity, omitted variables, or measurement errors. Due to the observation-driven nature of score models, the filtering method is simple and practical to implement. We establish the asymptotic properties of the maximum likelihood estimator and show that the instrumental-variable score-driven filter converges to the unique unknown causal path of the true parameter. We further analyze the finite sample properties of the filtered causal parameter in a comprehensive Monte Carlo exercise. Finally, we reveal the empirical relevance of this method in an application to aggregate consumption in macroeconomic data.

Keywords: causal inference, endogeneity, instrumental variables, observation-driven models, score model, time-varying parameters

*Corresponding author: n.stegehuis@vu.nl. This paper has been presented at the 2023 International Association for Applied Econometrics Annual Conference in Oslo (June 2023) and the International Conference on Econometrics and Statistics in Tokyo (August 2023). The authors thank participants at those conferences, as well as Paolo Gorgi, Siem Jan Koopman, Andre Lucas and Janneke van Brummelen for helpful comments and suggestions. We are grateful to Vipul Bhatt, N. Kundan Kishor and Hardik Marfatia for providing us with the data for the application. Francisco Blasques and Noah Stegehuis acknowledge the financial support of the Dutch Science Foundation (NWO) under grant Vidi.195.099.

1 Introduction

Establishing causal relationships between relevant variables is fundamental in economics and other fields of science. For example, causal inference is key for understanding the effects of fiscal and monetary policies in macroeconomics. It is however well known that standard predictive methods used in the econometric, statistical and machine learning literature will typically fail to uncover causal relationships when dealing with observational data, due to regressor endogeneity. As a result, considerable effort has been made to develop new and effective causal inference techniques. This effort, carried over the last few decades, as been recently recognized by the 2021 Nobel Memorial Prize in Economic Sciences awarded to Joshua Angrist and Guido Imbens. In particular, a range of solutions have been proposed in the literature for causal inference of observational data that typically suffers from endogeneity issues. Important examples include the use of instrumental variables (IV) (Angrist, Imbens, & Rubin, 1996), difference in differences (Ashenfelter & Card, 1985; Bertrand, Duflo, & Mullainathan, 2004) and regression discontinuity design (Thistlethwaite & Campbell, 1960) among others (e.g. synthetic control (Abadie & Gardeazabal, 2003), propensity score matching (Rosenbaum & Rubin, 1983)).

When dealing with time-series or panel data, it is important to note that causal inference methods generally attempt to uncover causal relationships that are assumed to be *time-invariant*. They are not designed to keep track of time-varying causal relationships, and do not focus on modeling that time-variation or producing dynamic forecasts of future cause-and-effect interactions. This can be, of course, a shortcoming in a number of applications. Much like in other scientific domains, in economics causal relationships can change over time both qualitatively and quantitatively. For example, the effectiveness of different fiscal and monetary policies may change substantially over time as they depend on the historical political, social, economic, technological and institutional context. In many practical applications, parameters can thus be time-varying and the need may arise to filter such parameters in order to track the evolution of the true causal effect and potentially forecast it. In macroeconomics for example, several studies find that the causal effect of monetary policy actions changed over time (Boivin & Giannoni, 2006). Attempts to identify these time-varying effects of monetary policy on macroeconomic variables have been made by Koop, Leon-Gonzalez, and Strachan (2009) and Korobilis (2013) among others.

In this paper we propose a novel score-driven filtering method (Creal, Koopman, & Lucas, 2013; Harvey, 2013) featuring instrumental variables to estimate the time-varying parameter in a regression model, with observational data, which may suffer from endogeneity. We model the endogeneity with a control function approach (e.g. Heckman and Robb (1985); Wooldridge (2015)) if suitable instruments are available. Similar to IV methods in general, our proposed method can potentially handle endogeneity originating from multiple causes. Important ones include: (i) simultaneity, which emerges when two variables are contemporaneously causally linked to each other (Haavelmo, 1943; Kennan, 1989); (ii) omitted variables, which occurs when regressors are correlated with other relevant regressors that are omitted from the regression model (Wooldridge, 2009); and (iii) measurement errors in relevant regressors (Bound, Brown, & Mathiowetz, 2001). Each of these conditions, and several others, result in endogeneity which renders predictive models unsuitable as tools for distilling causal effects (Haavelmo, 1943; Wooldridge, 2002).

Alternative regression models with time-varying parameters and endogenous regressors have been proposed in the literature. Kim (2006) and Kim and Kim (2011) propose both one-step and two-step estimation procedures of the Kalman Filter with a similar control function approach to handle endogeneity. Another approach that has recently been proposed by Giraitis, Kapetanios, and Marcellino (2021) uses a kernel based technique to estimate a time-varying IV estimator. Inoue, Rossi, and Wang (2022) propose a time-varying IV framework for Local Projections, based on the work of Müller and Petalas (2010), who show that for nonlinear non-Gaussian parameter-driven models with moderate time-variation, the sample information can be approximated by a linear Gaussian model that contains the scores of the likelihood as observations. This fact further motivates our use of score functions in the filter that estimates the time-varying parameter. The difference in performance between the score-driven filters and the method proposed by Müller and Petalas (2010) has been analysed in Calvori, Creal, Koopman, and Lucas (2017) in the context of a time-varying parameter test.

Our score-driven filter stands out in its simplicity of implementation and ability to produce robust and reliable path estimates in nonlinear non-Gaussian settings. Compared to parameter-driven models, such as the Kalman filter, observation-driven models like score-driven models are easier to implement and computationally less demanding, especially in nonlinear non-Gaussian parameter settings (Koopman, Lucas, & Scharth, 2016). Compared to non-parametric kernel methods, our parametric score-driven approach stands out in small-sample problems and forecasting exercises. Naturally, non-parametric methods can have advantages in terms of flexibility, but they will also require a choice of hyperparameters such as kernel related bandwidths.

Our filter fits more generally in the class of (quasi) score driven filters, as introduced by Creal et al. (2013), Harvey (2013) and Blasques, Francq, and Laurent (2023). The score driven method has been shown to have optimality properties over other methods such as the Kalman Filter. In particular, the updating scheme for a time-varying parameter is optimal in the information theoretic sense if and only if it contains the score of the likelihood; Blasques, Koopman, and Lucas (2015), Blasques, Koopman, and Lucas (2018), Blasques, Lucas, and van Vlodrop (2021), Beutner, Lin, and Lucas (2023). A similar optimality result holds for an analogous time-varying parameter approach in the context of the general method of moments (GMM) framework (Creal, Koopman, Lucas, & Zamojski, 2018). In the context of time-varying regression models, the score approach has been used by Blasques, Koopman, and Lucas (2020), Gorgi, Koopman, and Schaumburg (2017), in which autoregressive frameworks are considered. The time-varying parameter regression model with a score driven update is investigated by Thiele and Harvey (2013), who model time-varying correlations and by Blasques, Francq, and Laurent (2022), who introduce a time-varying beta model with GARCH dynamics for financial data. Notably however, none of these papers address the case of regressor endogeneity.

The rest of the paper is structured as follows. In Section 2 we describe the model and introduce the Instrumental Variables Score filter (IV-score). In Section 3 we analyse the stochastic properties of the filter. In Section 4 we show consistency of the two-step Maximum Likelihood Estimator (MLE) and show that the filter with estimated parameters converges to the unique path of the true unknown causal parameter when endogeneity is present. In a Monte Carlo simulation study in Section 5 we analyze the finite sample properties for various types and levels of endogeneity and we observe that the filter manages to uncover the true path even in non-stationary settings. In Section 6 we demonstrate the empirical relevance by applying the filter to estimate the excess sensitivity of consumption to income in the United States, as well as providing time-varying estimates of price elasticity of demand at the Fulton fish market.

2 Causal score-driven filtering model

Let $\{y_t\}_{t\in\mathbb{Z}}$ be a time series generated according to

$$y_t = \beta_t x_t + \varepsilon_t, \tag{1}$$

where $\{x_t\}_{t\in\mathbb{Z}}$ is a stochastic regressor, $\{\beta_t\}_{t\in\mathbb{Z}}$ is a time-varying parameter, and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a mean-zero identically distributed error term with density p_ε, indexed by a static parameter vector λ. We assume the regressors are endogenous, hence the usual exogeneity condition fails $\mathbb{E}(\varepsilon_t|x_t) \neq 0$. The specific cause of endogeneity can be left unspecified, but it could emerge from the usual culprits, ranging from simultaneity between y_t and x_t, omitted variables which are correlated with the regressor x_t, a measurement error in the regressor x_t, functional form misspecification, etc.

In the (quasi) score-driven approach, the filter for the time-varying parameter β_t is given by

$$\beta_{t+1} = \omega + \alpha s_t + \gamma \beta_t, \tag{2}$$

for fixed unknown parameters ω, α, γ, and where s_t is the scaled score,

$$s_t = S_t \cdot \nabla_t \qquad \nabla_t = \frac{\partial \ln p_y(y_t|x_t, \beta_t; \boldsymbol{\lambda})}{\partial \beta_t}, \tag{3}$$

with S_t being a scaling matrix and $p_y(y_t|\beta_t, x_t; \lambda) = p_\varepsilon(\tilde{\varepsilon}_t; \lambda)$, where $\tilde{\varepsilon}_t = y_t - \beta_t x_t$ is the prediction error. For example, for normally distributed homoskedastic disturbances

$p_\varepsilon(\varepsilon_t;\lambda) = f(\varepsilon_t;\sigma_\varepsilon^2)$, where $f(\cdot)$ is the density function of a mean-zero Gaussian distribution with variance σ_ε^2, we get $\nabla_t = x_t(y_t - \beta_t x_t)\sigma_\varepsilon^{-2}$, giving rise to the filtering equation

$$\beta_{t+1} = \omega + \alpha\sigma_\varepsilon^{-2} x_t(y_t - \beta_t x_t) + \gamma\beta_t. \tag{4}$$

Intuitively, this specification of the score ensures that the parameter will be updated to reduce the latest prediction error. In this way the parameter stays up-to-date and provides the closest model fit based on the most recent observations. Clearly, other distributions for the error term can be chosen and the parameter update will naturally be adjusted.

When regressors are exogenous, the static parameters of the filtering equation in (2), collected in the vector $\boldsymbol{\theta} = (\omega, \alpha, \gamma, \sigma_\varepsilon^2)'$, can be consistently estimated by maximum likelihood, and ultimately, a time-varying β_t can be adequately filtered; see e.g. Blasques, Gorgi, Koopman, and Wintenberger (2018) for filter convergence results. This means essentially that the filtered parameter $\hat{\beta}_t(\hat{\boldsymbol{\theta}}_T)$ initialized at some value $\hat{\beta}_1$ in a correctly specified model, will converge to its true unobserved value β_t as both t and T diverge to infinity,

$$|\hat{\beta}_t(\hat{\boldsymbol{\theta}}_T, \hat{\beta}_1) - \beta_t| \xrightarrow{p} 0 \quad \text{as } (t,T) \to \infty. \tag{5}$$

Taking t to infinity is required so that the effect of the (almost surely) incorrect initialization of the filter at $\hat{\beta}_1$ vanishes. This is ensured by establishing the so-called *invertibility* of the filter, which requires the filter to have fading memory. Taking the sample size T to infinity ensures that the MLE of the static parameters $\hat{\boldsymbol{\theta}}_T$ converges to $\boldsymbol{\theta}_0$.

Unfortunately, this same score filtering technique will fail to uncover the causal β_t when the regressor x_t is endogenous. Indeed, when the exogeneity condition $\mathbb{E}[\varepsilon_t|x_t] \neq 0$ fails due to simultaneity, omitted variables, or other factors, the filter convergence in (5) will no longer hold. In order to handle the problem of regressor endogeneity, we take a two-step instrumental variable approach. This method requires the existence of valid instruments $\{z_t\}_{t\in\mathbb{Z}}$, that are used to estimate the first stage regression

$$x_t = \pi z_t + u_t \qquad\qquad u_t \overset{i.i.d}{\sim} N(0, \sigma_u^2), \tag{6}$$

where $\mathbb{E}[u_t|z_t] = 0$ and the true parameter $\pi_0 \neq 0$. In a second step, we take a control function (CF) approach to correct for endogeneity. Like in a static model (see Wooldridge, 2015), the correlation between the regressor and error term in (1) can be modeled as $\varepsilon_t = \tau u_t + \eta_t$, where we let η_t is an identically and independently distributed (i.i.d.) innovation regardless of the type of endogeneity. As such, the relation between y_t and x_t is given by $y_t = \beta_t x_t + \tau u_t + \eta_t$. Analogous to the static control function approach, we add the first stage fitted residuals to the structural equation to replace the unknown first stage errors. As a result, after obtaining the residuals $\hat{u}_t = x_t - \hat{\pi}_{T_1} z_t$ from the first-stage regression, where $\hat{\pi}_{T_1}$ denotes the estimate of π. To be explicit about the fact that the first stage estimate is carried over to the second stage, we denote the sample sizes of both stages T_1 and T_2 respectively. The second stage regression model is given by

$$y_t = \beta_t x_t + \tau \hat{u}_t + \eta_t. \tag{7}$$

Further, the Gaussian causal IV-score filtering equation that assumes $\eta_t \overset{i.i.d}{\sim} N(0, \sigma^2)$, takes the form

$$\beta_{t+1} = \omega + \alpha\sigma^{-2} x_t (y_t - \beta_t x_t - \tau(x_t - \hat{\pi}_{T_1} z_t)) + \gamma\beta_t \tag{8}$$

where $\hat{\pi}_T$ has been estimated ex ante through equation (6), and we collect the static parameters in $\boldsymbol{\theta} = (\omega, \alpha, \gamma, \tau, \sigma^2)'$.

This control function approach bears similarity to Terza, Basu, and Rathouz (2008), who highlight the importance of using a CF in nonlinear linear-index models to avoid inconsistent estimation. Note that simply replacing the regressors with the fitted values of the first stage in the filter in (4), a more common 2SLS approach, will not work. In time-invariant models, the knowledge that the prediction error is zero on average and that the loss function should be minimized is enough. But in this setting, next to its direct role in loss function minimization, the prediction error also drives the filter, hence a good estimate for it is crucial. It is necessary therefore to control for the movements in the regressors as much as possible, by adding the first stage residuals to the structural equation. Not doing that will result in the prediction error, and along with it the filtered path, being governed by unobserved movements in x_t. In the case of a highly relevant but omitted variable for example, the prediction error and filter will then mimic the omitted variable rather than distill the causal effect.

We note that all of the subsequent results extend easily to the case in which models defined in (1), (6) and (7) have an intercept $a \in \mathbb{R}$ so that $y_t = a + \beta_t x_t + \eta_t$. For simplicity, we let $a = 0$ and assume that the data is demeaned. We note that the current model can also be easily extended to allow for a time-varying parameter π_t, by constructing a multivariate filter similar to Blasques, Francq, and Laurent (2022). For theoretical simplicity we continue using a static π as we focus on the causal time-varying parameter of interest β_t, and leave this extension for future research.

3 Stochastic Properties of the Filter

We collect all data at time t in the vector $Y_t := (y_t, x_t, z_t)'$. We denote the sample history of this vector by $Y^{1:t} := \{Y_1, \ldots, Y_{t-1}, Y_t\}$, and the entire history of this vector stretching to the infinite past by $Y^t := \{\ldots, Y_{t-1}, Y_t\}$. We start from the filter definition in (2) to allow for a flexible framework that includes more general models with different error distributions and model specifications, among others. We denote the filtered sequence, that depends on the sample data, by $\{\hat{\beta}_t(Y^{1:t-1}, \boldsymbol{\theta}, \pi, \hat{\beta}_1)\}_{t\in\mathbb{N}}$, with short hand $\hat{\beta}_t(\boldsymbol{\theta}, \pi, \hat{\beta}_1)$

also sometimes denoted by $\hat{\beta}_t$. The filtered sequence initialized in the infinite past is denoted by $\{\tilde{\beta}_t(Y^{t-1}, \boldsymbol{\theta}, \pi)\}_{t\in\mathbb{Z}}$ or simply $\tilde{\beta}_t := \tilde{\beta}_t(\boldsymbol{\theta}, \pi)$. If the model is correctly specified and both $\boldsymbol{\theta}_0$ and π_0 are the true parameters, then $\beta_t^o := \tilde{\beta}_t(\boldsymbol{\theta}_0, \pi_0)$ is the true stochastic time-varying parameter. We let $\boldsymbol{\lambda}$ contain all model and density specific parameters (in equation (8) $\boldsymbol{\lambda} = (\tau, \sigma^2)'$) and let $\boldsymbol{\theta} = (\omega, \alpha, \gamma, \boldsymbol{\lambda})$ and $k = \dim(\boldsymbol{\theta})$. For a random variable $x(\boldsymbol{\theta})$ possibly depending on $\boldsymbol{\theta} \in \Theta$, we further let $\|x(\cdot)\|_n^{\Theta} := (\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |x(\boldsymbol{\theta})|^n)^{1/n}$ and $\|x(\cdot)\|_{\Theta} := \sup_{\boldsymbol{\theta}\in\Theta} |x(\boldsymbol{\theta})|$. Finally, we define the stochastic function

$$\Lambda_t^*(\boldsymbol{\theta}, \boldsymbol{\theta}^*, \pi) := \sup_{\beta^* \in \mathcal{F}_{\boldsymbol{\theta}^*}} |\gamma + \alpha \partial s(\beta, Y_t, \pi; \boldsymbol{\theta})/\partial\beta|_{\beta=\beta^*}| \tag{9}$$

for some $\pi \in \mathbb{R}$. Then $\Lambda_t(\boldsymbol{\theta}, \pi) := \Lambda_t^*(\boldsymbol{\theta}, \boldsymbol{\theta}, \pi)$ is the special case where $\boldsymbol{\theta}^* = \boldsymbol{\theta}$.

Invertibility

Lemma 1 gives general conditions for invertibility of the filter. Invertibility ensures that the effect of the initialization of the filter vanishes in the limit, meaning that the filtered sequence converges to its unique limit for any given initialization $\hat{\beta}_1$. Lemma 1 further establishes stationarity and ergodicity (SE) of the limiting sequence, which will be used when deriving the asymptotic properties of the maximum likelihood estimator. Note that this result holds irrespective of whether the model is correctly specified.

Lemma 1 (Invertibility)**.** *Let $\Theta \subset \mathbb{R}^k$ be compact, and let the elements in $\{Y_t\}_{t\in\mathbb{Z}}$ be SE sequences. Let $\hat{\pi}_{T_1} \in \mathbb{R}$ and assume there exists some $\hat{\beta}_1 \in \mathcal{F}$ such that*

(i) $\mathbb{E}\log^+ \sup_{\boldsymbol{\theta}\in\Theta} |s(\hat{\beta}_1, Y_t, \hat{\pi}_{T_1}; \boldsymbol{\theta})| < \infty$

(ii) $\mathbb{E}\log \sup_{\boldsymbol{\theta}\in\Theta} \Lambda_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}) < 0$

Then the sequence $\{\hat{\beta}_t\}_{t\in\mathbb{N}}$ converges exponentially almost surely to a unique limit SE sequence $\{\tilde{\beta}_t\}_{t\in\mathbb{Z}}$ uniformly on Θ , i.e. $\|\hat{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_1) - \tilde{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1})\|_{\Theta} \xrightarrow{e.a.s} 0$ as $t \to \infty$.

In Corollary 1 we formulate the invertibility conditions for the specific case of the Gaussian IV-score filter in equation (8). The conditions for the score filter with exogenous regressors in equation (4) follow with $\tau = 0$. The invertibility conditions (i) and (ii) are trivially satisfied for the Gaussian IV-score filter.

Corollary 1 (Invertibility for Gaussian IV-score)**.** *Let $\{Y_t\}_{t\in\mathbb{Z}}$ be a SE sequence and let $\hat{\pi}_{T_1}$ be the first stage estimate of π_0. Let $\boldsymbol{\theta} = (\omega, \alpha, \gamma, \tau, \sigma^2)' \in \Theta$ where $\Theta \subset \mathbb{R}^5$ is compact and assume there exists some $\hat{\beta}_1 \in \mathcal{F}$ such that*

(i) $\mathbb{E}\log^+ \sup_{\boldsymbol{\theta}\in\Theta} |\sigma^{-2} x_t(y_t - \hat{\beta}_1 x_t - \tau(x_t - \hat{\pi}_{T_1} z_t))| < \infty$

(ii) $\mathbb{E}\log\sup_{\boldsymbol{\theta}\in\Theta}|\gamma-\alpha\sigma^{-2}x_t^2|<0$

Then the sequence $\{\hat{\beta}_t\}_{t\in\mathbb{N}}$ *produced by the filtering equation* (8) *converges e.a.s. to a unique limit SE sequence* $\{\tilde{\beta}_t\}_{t\in\mathbb{Z}}$ *uniformly on* Θ *, i.e.* $\|\hat{\beta}_t(\boldsymbol{\theta},\hat{\pi}_{T_1},\hat{\beta}_1)-\tilde{\beta}_t(\boldsymbol{\theta},\hat{\pi}_{T_1})\|_\Theta \xrightarrow{e.a.s} 0$ *as* $t\to\infty$.

Bounded moments

As we shall see in Section 4, beyond filter invertibility, the MLE consistency proof that we establish will also require that the limit filter has bounded moments when evaluated at the true parameter $(\pi_0,\boldsymbol{\theta}_0)\in\mathbb{R}\times\Theta$. According to Lemma 1, the limit filter evaluated at $(\pi_0,\boldsymbol{\theta}_0)$, satisfying the recurrence $\beta_{t+1}=\omega_0+\alpha_0\sigma_0^{-2}x_t(y_t-\beta_t x_t-\tau_0(x_t-\pi_0 z_t))+\gamma_0\beta_t$, converges to the same unique solution

$$\beta_{t+1}=\omega_0+\alpha_0\sigma_0^{-2}x_t\eta_t+\gamma_0\beta_t. \tag{10}$$

This is a stochastic recurrence equation of the type $\beta_{t+1}=\phi_0(x_t,\eta_t,\beta_t)$ with derivative given by $\partial\phi_0(x_t,\eta_t,\beta_t)/\partial\beta=\gamma_0$ where ϕ_0 is a function defined by the parameter vector $(\pi_0,\boldsymbol{\theta}_0)$. Lemma 2 and Corollary 2 establish bounded moments for this limit process.

Lemma 2 (Limit filter moments)**.** *Let* $\{Y_t\}_{t\in\mathbb{Z}}$ *be a SE sequence. Suppose* $\exists\, n_\beta>0$ *such that* $\|\phi_0(x_t,\eta_t,\beta_t)\|_{n_\beta}<\infty$, *and* $\sup_{(\beta^*,Y)\in\mathcal{F}\times\mathcal{Y}}|\partial\phi_0(x_t,\eta_t,\beta_t)/\partial\beta|<1$. *Then* $\|\tilde{\beta}_t(\boldsymbol{\theta}_0,\pi_0)\|_{n_\beta}<\infty$.

Corollary 2 (Limit filter moments for Gaussian IV-score)**.** *Let* $\{Y_t\}_{t\in\mathbb{Z}}$ *be a SE sequence. Suppose* $\exists\, n_\beta>0$ *such that* $(\mathbb{E}|x_t\eta_t|^{n_\beta})^{1/n_\beta}<\infty$ *and* $|\gamma_0|<1$. *Then the limit sequence* $\|\tilde{\beta}_t(\boldsymbol{\theta}_0,\pi_0)\|_{n_\beta}<\infty$.

4 Asymptotic Properties of the Maximum Likelihood Estimator

An advantage of the score-driven time-varying parameter models is that the static parameter $\boldsymbol{\theta}_0\in\Theta$ can be estimated by a straight forward maximum likelihood (ML) procedure. We define the second stage ML estimator as

$$\hat{\boldsymbol{\theta}}_{T_2}(\pi)\in\arg\max_{\boldsymbol{\theta}\in\Theta}\ell_{T_2}(\boldsymbol{\theta},\pi,\hat{\beta}_1) \tag{11}$$

where

$$\ell_{T_2}(\boldsymbol{\theta},\pi,\hat{\beta}_1)=\frac{1}{T_2}\sum_{t=1}^{T_2}\ell_t(\boldsymbol{\theta},\pi,\hat{\beta}_t(\boldsymbol{\theta},\pi,\hat{\beta}_1)) \tag{12}$$

and $\ell_t(\boldsymbol{\theta}, \pi, \hat{\beta}_t(\boldsymbol{\theta}, \pi, \hat{\beta}_1) := \log p_y(y_t|\hat{\beta}_t(\boldsymbol{\theta}, \pi, \hat{\beta}_1), x_t, z_t, \boldsymbol{\theta}, \pi)$. Furthermore, define $\ell_0(\boldsymbol{\theta}, \pi_0) := \ell_t(\boldsymbol{\theta}, \pi_0, \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0))$ and $\ell_\infty(\boldsymbol{\theta}, \pi_0) = \mathbb{E}[\ell_0(\boldsymbol{\theta}, \pi_0)]$. To establish consistency in similar spirit to Blasques, Gorgi, et al. (2018), assume that the following conditions hold:

(C1) The DGP which satisfies equations (6) to (8) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0 \in \Theta$ admits a stationary solution and that $\{(x_t, z_t)\}_{t\in\mathbb{Z}}$ is a SE sequence.

(C2) $\mathbb{E}|\ell_0(\boldsymbol{\theta}_0, \pi_0)| < \infty$

(C3) For any $\boldsymbol{\theta} \in \Theta$, $\ell_0(\boldsymbol{\theta}_0, \pi_0) = \ell_0(\boldsymbol{\theta}, \pi_0)$ if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$

(C4) The invertibility conditions (i) and (ii) of Lemma 1 are satisfied for the compact set $\Theta \subset \mathbb{R}^k$

(C5) The sequences $\left\{ \left\| \frac{\partial \ell_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \beta)}{\partial \beta} \Big|_{\beta = \tilde{\beta}_t^*(\boldsymbol{\theta}, \pi)} \right\|_\Theta \right\}_{t\in\mathbb{Z}}$ and $\left\{ \left\| \frac{\partial \ell_t(\boldsymbol{\theta}, \pi, \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0))}{\partial \pi} \Big|_{\pi = \pi_{T_1}^*} \right\|_\Theta \right\}_{t\in\mathbb{Z}}$ are SE, where $\tilde{\beta}_t^*(\boldsymbol{\theta}, \pi)$ is a point between $\tilde{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1})$ and $\tilde{\beta}_t(\boldsymbol{\theta}, \pi_0)$ and $\pi_{T_1}^*$ is a point between $\hat{\pi}_{T_1}$ and π_0.

(C6) $E\|\ell_0(\boldsymbol{\theta}, \pi_0) \vee c\|_\Theta < \infty$ for some $c \in \mathbb{R}$ such that $c < \ell_\infty(\boldsymbol{\theta}_0, \pi_0)$

(C7) $\hat{\pi}_{T_1} \xrightarrow{p} \pi_0$ as $T_1 \to \infty$

As noted in Blasques, Gorgi, et al. (2018), we obtain strong consistency following Wald (1949). Theorem 1 states that the maximum likelihood estimator evaluated at some initialization $\hat{\beta}_1$ and at the first stage estimate $\hat{\pi}_T$ is consistent for the true unknown parameter. Corollary 3 provides the same consistency result but applied to the special case of our Gaussian IV-score model.

Theorem 1 (Consistency)**.** *Let the conditions* (C1)-(C7) *hold. Then the maximum likelihood estimator is consistent*

$$\hat{\boldsymbol{\theta}}_{T_2}(\hat{\pi}_{T_1}, \hat{\beta}_1) \xrightarrow{p} \boldsymbol{\theta}_0 \qquad T_1, T_2 \to \infty, \tag{13}$$

for any initialization $\hat{\beta}_1 \in \mathcal{F}$*.*

Corollary 3 (Consistency MLE of Gaussian IV-Score)**.** *Let the process* $\{y_t\}_{t\in\mathbb{Z}}$ *be generated by the model in equations equations* (6) *to* (8) *with* $\boldsymbol{\theta} = \boldsymbol{\theta}_0 \in \Theta$ *such that* $|\gamma_0| < 1$ *and* $\sigma_0^2 > 0$*, and let the sequences* $\{(x_t, z_t)\}_{t\in\mathbb{Z}}$ *be SE. Furthermore let* Θ *be compact such that* $\mathbb{E}\log|\gamma - \alpha\sigma^{-2}x_t^2| < 0$ *and* $\sigma > 0$ $\forall \boldsymbol{\theta} \in \Theta$*. Then the maximum likelihood estimator* $\hat{\boldsymbol{\theta}}_{T_2}(\hat{\pi}_{T_1}, \hat{\beta}_1)$ *with any initialization* $\hat{\beta}_1 \in \mathcal{F}$ *is consistent.*

Building on the consistency of the MLE, we can also provide a convergence result of the causal time-varying parameter of interest. Proposition 1 assures that the filter converges to the true time-varying parameter, and is a direct result of the previously established invertibility of the filter and consistency of the MLE. Strong consistency in part (b) holds whenever the limit process $\{\tilde{\beta}_t\}_{t\in\mathbb{Z}}$ has a log plus bounded moment. Although this is a rather mild condition, that will most likely hold due to the stationarity of this limit process, but cannot always be confirmed theoretically[1].

Proposition 1 (Path Convergence)**.** *Let* (C1) *and* (C4) *hold and* $\hat{\boldsymbol{\theta}}_{T_2} \xrightarrow{p} \boldsymbol{\theta}_0$*. Then the IV-score filter is consistent for the true unobserved causal time-varying parameter* $\{\beta_t^o\}_{t\in\mathbb{Z}}$ *when evaluated at* $\hat{\pi}_{T_1}$*,*

$$|\hat{\beta}_t(\hat{\boldsymbol{\theta}}_{T_2}, \hat{\pi}_{T_1}, \hat{\beta}_1) - \beta_t^o| \xrightarrow{p} 0 \text{ as } T_2 \geq T_1 \geq t \to \infty \tag{14}$$

for any initialization $\hat{\beta}_1 \in \mathcal{F}$*.*

5 Simulation Study

In the following simulation study we investigate the performance of the new IV-score filter compared to a regular score filter, as specified in (8) and (4) respectively, for a time-varying parameter regression model in which the regressor is endogenous. In terms of endogeneity, we consider the general formulation that encompasses all types of endogeneity, among which omitted variables, measurement errors and simultaneity.

We consider the following data generating process (DGP1), in which the true time-varying parameter β_t as well as the instrument z_t, are generated by stationary AR(1) processes for given parameters $\tau, \pi, \sigma_\eta, \sigma_u, \sigma_\beta, \sigma_z$.

$$\begin{aligned}
y_t &= \beta_t x_t + \tau u_t + \eta_t & \eta_t &\sim N(0, \sigma_\eta^2), \\
x_t &= \pi z_t + u_t & u_t &\sim N(0, \sigma_u^2), \\
\beta_t &= 0.1 + 0.95\beta_{t-1} + \xi_t & \xi_t &\sim N(0, \sigma_\beta^2), \\
z_t &= 0.2 z_{t-1} + \zeta_t & \zeta_t &\sim N(0, \sigma_z^2).
\end{aligned}$$

We generate a path of $T_1 = T_2 = T = 1000$ observations for $\{\beta_t\}_{t=1}^T$ once, and subsequently draw the data $M = 1000$ times, estimate both filters and evaluate them by taking the

[1]Note that in Lemma 2 we established bounded moments for the limit filter when evaluated at the true parameter $(\tau_0, \boldsymbol{\theta}_0)$, which conveniently changes the prediction error into exactly η_t. Proposition 1 (b) requires a log plus bounded moment uniformly over $\boldsymbol{\theta}$, which for the Gaussian IV-score filter for example, results in the condition $\mathbb{E}\sup_{(\beta^*,Y,\boldsymbol{\theta})\in\mathcal{F}\times\mathcal{Y}\times\Theta}|\gamma - \alpha\sigma^{-2}x_t^2| < 1$. This condition is not evident to hold without further assumptions on x_t, since the supremum is taken over the data.

average mean squared error (MSE) over the whole path. With the parameters τ, π, σ_u we can increase and decrease the level of endogeneity and determine the strength of the instrument, while we fix $\sigma_\eta^2 = 1, \sigma_\beta^2 = 0.1, \sigma_z^2 = (1 - 0.2^2)$. The true path for β_t is chosen with a high persistency parameter, as real-life processes most likely evolve slowly over time.

Uncovering the True Causal Parameter

In Figure 1 and Table 1 we present the results for DGP 1 for the case where $\tau = -4, \pi = 1$ and $\sigma_u = 5$. This is a configuration of the parameters for which the bias is substantial and the illustration is clear. The plots of the paths are constructed by taking medians at each point $t = 1, \ldots, T$ over all simulations. In the figures we include the true path (blue), the static least squares estimator (OLS, red) and IV (green) estimators as estimated on the whole sample, and the median IV-score and score (orange) paths over all simulations.

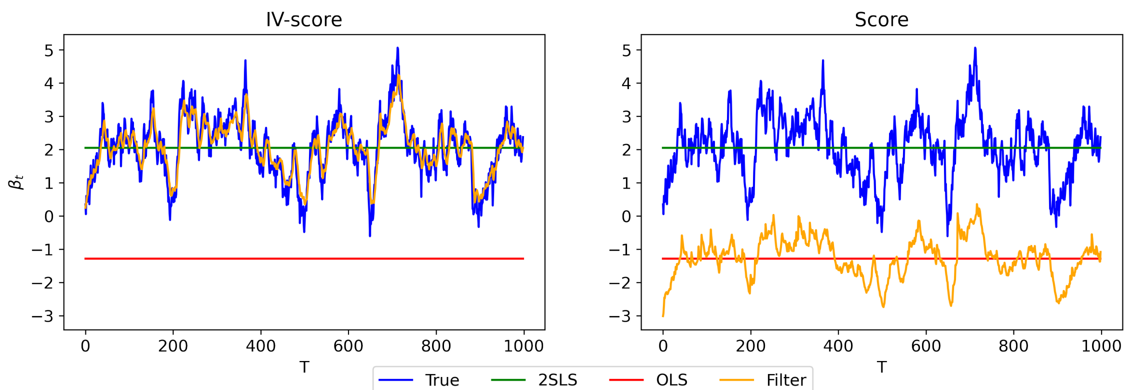

Figure 1: Estimated paths of the causal parameter with DGP 1 where β_t follows an AR(1).

	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\gamma}$	$\hat{\beta}_0$	$\hat{\sigma}_\eta^2$	$\hat{\tau}$	b_0	b_1	MSE
OLS							0.129	-1.635	12.109
IV						-3.740	0.022	1.762	0.974
Score	-0.062	0.353	0.952	-3.289	16.992				11.663
IV-score	0.128	0.121	0.946	0.409	2.448	-4.263			**0.349**

Table 1: Maximum Likelihood Estimates and MSE (DGP 1).

Comparing the MSEs of the filtered paths relative to the true causal paths (Table 1) it becomes clear that the IV-score filter outperforms the score filter without difficulty,

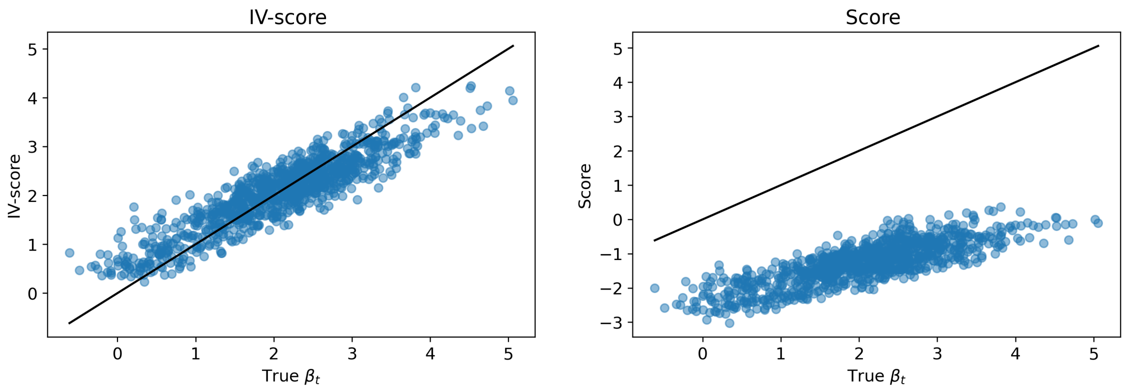

Figure 2: Scatter plot of IV-score and score filtered values vs. true β_t (DGP 1).

highlighting that taking endogeneity into account is crucial. The endogeneity bias is visible in Figure 1 through the large gap between the static OLS and IV estimators. The OLS estimator suggests a consistent negative relation, while much of the causal parameter is actually positive, something that is correctly picked up on by the static IV estimator. This bias in the OLS estimator is carried over to the score-driven model in the time-varying case. In contrast to this regular score-driven model, the IV-score filter captures the true causal relation well.

At a first glance it seems that the aforementioned static OLS-IV bias shift is the only difference between the score and IV-score filtered paths. But not only is it shifted, the score filter does not follow the same dynamics as the true parameter does. This is clearly shown in the scatter plots of the true parameter versus the filtered parameters displayed in Figure 2. The IV-score filter moves along with the true parameter in the right directions (high when high), while the score filter produces a much flatter relation. Note that for none of the (subsequent) DGPs we simulate the true time-varying parameter to follow our IV-score specification, but let the process evolve independently. Another crucial observation is that this DGP is parameter-driven, while our (IV)score method is observation-driven. This shows that misspecification in this sense does not affect the performance of our observation-driven approach. For more discussion on this comparison, see Koopman et al. (2016).

Robustness Check

In this section we investigate the robustness of the method. In particular, we explore whether the method breaks down with (i) a non-stationary causal parameter (DGP 2 & 3), (ii) when the instrument is not fully exogenous (DGP 4), (iii) when the observations have high variance (DGP 5) and (iv) when the data contains outliers (DGP 6). The DGP's in each of these four scenarios are based on DGP 1 with the following adaptations.

DGP 2	$\beta_t = \begin{cases} -1 & t \leq \lfloor T/2 \rfloor \\ 1 & t > \lfloor T/2 \rfloor \end{cases}$
DGP 3	$\beta_t = 1 + \beta_{t-1} + \xi_t$
DGP 4	$z_t = 0.2 z_{t-1} + 0.2 u_t + \zeta_t$
DGP 5	$\sigma_\eta^2 = 25$
DGP 6	$y_t = \beta_t x_t + \tau u_t + (1 - I_t)\eta_t + I_t S_t \phi_t$
	$I_t \sim \text{Bernoulli}(0.01), S_t \sim \text{Uniform}\{-1, 1\}, \phi_t \sim N(50, 5^2)$

Non-stationary Causal Parameter
Figure 6 in Appendix B shows the filtered path for a midway break in β_t generated by DGP 2, jumping from a negative value to positive in the middle of the sample. Even in this non-stationary setting, the IV-score filter manages to capture the change in the parameter very closely, with only little variation around the true line. The score filter on the other hand does show some indication of a break in levels, but remains negative throughout the whole sample due to the bias. When we generate a more severe case of non-stationarity by letting β_t follow a random walk (DGP 3 and Figure 7), we still find that the filter performs considerably well.

Contaminated Instrument
When the instrument is not exogenous but also correlated with the error term u_t (DGP 4), the IV-score filter breaks down as expected. In Figure 8 the paths are both visibly biased, although the IV-score filter still benefits from some exogenous movements in z_t to get closer the the true path. This figure clearly illustrates how the performance depends on how the static IV estimator with the same endogeneous instruments relates to the true path. The lower the bias of IV, the better the performance of the related IV-score filter.

Large Error Variance
When the variance of the error term is increased to $\sigma_\eta^2 = 25$, the result is a more flattened filter as visible in Figure 9. This is to be expected, as the prediction error in the score will have larger outlying values and the corresponding estimate $\hat{\alpha}\hat{\sigma}_\eta^{-2}$ will adjust accordingly making the overall filter less responsive to any changes in the score, also those caused by the true parameter. The result is a much less responsive filter that therefore mimics the true time-varying parameter more poorly.

Outliers
Outliers in the data might influence the filtered path, although it depends on their size and frequency. For example, simulating the errors from a fat tailed t-distribution with

$\nu = 3$ degrees of freedom, will not generate any sudden jumps in the filtered path, as the parameter α pre-multiplying the score will be adjusted towards zero. Therefore we examine with DGP 6 a situation in which there are only a few extreme errors, taking values around 50, in contrast to rest that are standard normally distributed. Although this leads to extreme values of the data, the outliers only slightly affect the filtered path, as the filter is somewhat flattened out by a smaller value of $\hat{\alpha}\hat{\sigma}_\eta^{-2}$. In Figure 10 we do observe some peaks where the outliers appear (gray lines), but it still provides an acceptable path given such extreme outliers.

6 Empirical Applications

Excess Sensitivity of Consumption

We now provide an illustration of our score-driven filter for endogenous regressors, by estimating time-varying excess sensitivity of consumption to income. Originally, Hall (1978) hypothesized that consumption (C_t) was a random walk, while later Campbell and Mankiw (1989) proposed an extended model by assuming that a fraction of the population (λ) consumes out of their current income (Y_t). As a result, a part of the changes in income drive changes in consumption, expressed as

$$\Delta C_t = \alpha + \lambda \Delta Y_t + \varepsilon_t.$$

Due to the potential endogeneity of income, Campbell and Mankiw (1989) suggested using lagged variables as instruments. It was only recently, that Bhatt, Kundan Kishor, and Marfatia (2020) highlighted that although these instruments have been argued to be exogenous, they are in fact weak instruments. Instead, these authors proposed to use the lagged 1-step-ahead Greenbook forecasts of changes in real disposable income, as these are simultaneously highly correlated with the real disposable income at time t (as opposed to the previously used instruments), and exogenous due to the forecasting nature of the variable, thus eliminating any undesirable contemporaneous effects.

The parameter of interest is λ, which measures the causal impact that changes in income have over changes in aggregate consumption. By sub-sample estimation (1978-1999 and 2000-2010), Bhatt et al. (2020) find that λ is much smaller in the most recent sub-sample than in the sample based on the years before 2000, which motivates their choice for taking a time-varying approach. Similarly, we estimate the time-varying parameter using our score-driven causal filter, while adding a lag of ΔC_t with a fixed parameter[2] to capture the effect of changes in income on consumption growth. The structural model

[2] Bhatt et al. (2020) also add ΔC_{t-1} to their model, however with a time-varying parameter ϕ_t. Such a multivariate filter is also possible in our framework, but for simplicity and brevity we take a constant parameter as we wish to merely illustrate the use of our filter for the parameter of interest λ_t.

equation we estimate is

$$\Delta C_t = \alpha + \lambda_t \Delta Y_t + \phi \Delta C_{t-1} + \varepsilon_t, \tag{15}$$

where we use the Greenbook forecasts (Y_t^{GB}) as instruments, and the model and filtering equation are analogous to equations (6), (7), (8).

Figure 3 (left) shows the growth rates of aggregate consumption, national income, as well as the GreenBook forecasts. Similar to Bhatt et al. (2020), we take quarterly observations of real consumption expenditures per capita on non-durable goods and services and real per capita disposable income from 1978-2010 in the U.S., as well as the the one quarter forecast of real disposable income from the Greenbook reports[3].

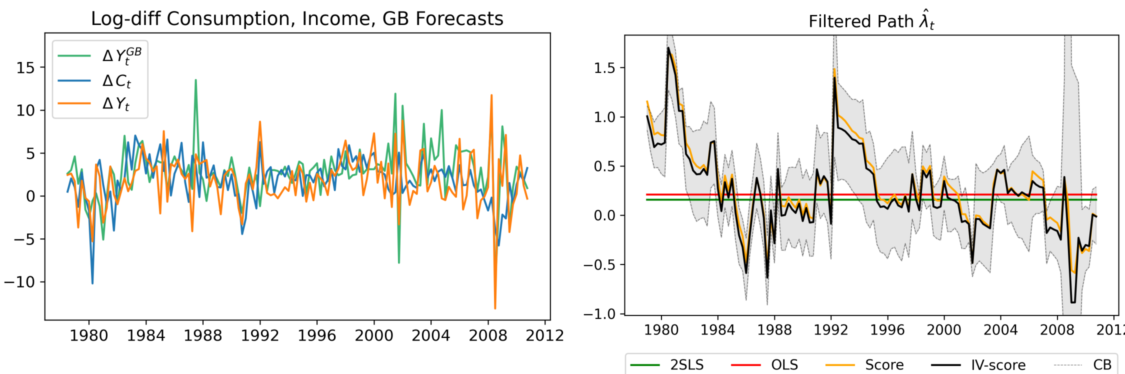

Figure 3: Data and filtered path of excess sensitivity parameter λ_t

In the right panel of Figure 3 we plot the static OLS and IV estimators and the resulting score-driven filtered paths of λ_t, accounting for endogeneity and not (IV-score and score respectively). The area shaded gray are the the 90% confidence bands for the IV-score filter, that are estimated using a bootstrap procedure accounting for parameter uncertainty, inspired by Pascual, Romo, and Ruiz (2006) and modified for our two-step score-driven model with $B = 500$ replications.

From the estimated filter paths we observe that sensitivity of consumption to income seems to have steadily decreased during the 80s. Over the next decade the sensitivity increased somewhat until the sudden drop in the wake of the 2001 crisis. There seems to have been speedy recovery however, until the plunge due to the 2008 financial crisis. The confidence bands of the IV-score filter suggest that there were several periods in which there was a significant effect, particularly when the filter takes on its extreme values. Moreover, the bands are wider during unpredictable times, as seen from the changing

[3]We are grateful to the authors Bhatt et al. (2020) for providing us with the data. For a detailed description of the source and type of data, see Bhatt et al. (2020).

distance between the lower and upper bound over time. This is most prominent during the global financial crisis, in which the bands even stretch in opposite directions, capturing the additional uncertainty of this event. The reason is an unusually large deviation in ΔY_t, which in our filter specification is multiplied with the prediction error, resulting in wider confidence bands that reflect the increased uncertainty in volatile periods.

An important observation is that endogeneity does not seem to be notably present as the static OLS and IV are not vastly different. This is indeed confirmed by a Hausman test with a p-value of 0.55. It is however possible that the level of endogeneity also varies over time, and that the Hausman statistic that only measures the average effect for static estimators, averages out any local or temporary endogeneity over the entire sample. In the absence of standard errors on the parameter estimates themselves, we cannot make any such conclusions based on the value of the first stage residual parameter estimate ($\hat{\tau} = 0.122$), but Bhatt et al. (2020) find in their time-varying model that it is significant, hence that endogeneity is in fact a concern. Comparing the IV-score and score paths, we indeed find that they overlap for a large part of the quarters in the sample, but they do differ substantially in a few instances.

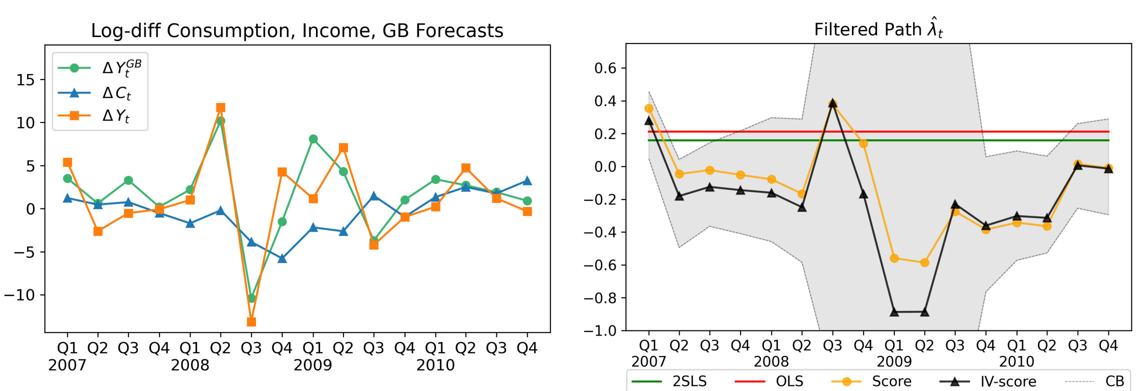

Figure 4: Detail global financial crisis: data and filtered path of λ_t

Most notably, during the period 2008Q4-2009Q2 right after the start of the global financial crisis with the bankruptcy of Lehman Brothers, the estimated effect by IV-score deviates from the regular score filter (see Figure 4 for a zoomed-in version of this period). In particular, we measure that the effect of income on consumption in that period turns out to be larger, in fact almost twice as large, once we account for endogeneity. This is broadly in line with the observation that during crises the sensitivity to income is often underestimated, as was pointed out by Blanchard and Leigh (2013), who reveal that fiscal multipliers were much higher than the initial forecasts during the great recession. The results are also in accordance with observations by Eggertsson and Krugman (2012), who suggest that the Ricardian equivalence broke down during this period, implying the reappearance of classic Keynesian type multipliers, in which current consumption depends more strongly on current income.

Fulton Fish Market

We provide a second empirical application in which the bias in the score path is more severe. We consider supply and demand models, that originally motivated Wright (1928) to introduce instrumental variables. In any such supply/demand setting, average prices and total quantities sold are recorded each period and one might be interested in the evolution of price elasticity of demand over time. However, endogeneity is present in the data, which stems from the fact that the observed prices and quantities are determined in equilibrium. One such case of price elasticity of demand has been investigated by Graddy (1995) and Angrist, Graddy, and Imbens (2000), at the New York Fulton fish market. In order to provide a consistent estimate of the causal effect of log price on log quantity sold of pounds of whiting fish, Angrist et al. (2000) use a weather variable as an instrument that represents how stormy the weather conditions were at sea the day before. The validity of this instrument follows from the fact that it shifts the supply curve without affecting demand.

Applying our score driven filter to this data, using the storm variable as an instrument, we arrive at the same result that the actual structural effect of price on quantity is double the effect compared to least squares estimates. Our time varying estimated paths generated by the score and IV-score filters, are similar in movements but fluctuate around these different levels of the OLS and IV estimators. The 90% confidence band that captures parameter uncertainty shows that the time-varying parameter is significant for most time periods, and is rather wide relative to the path estimate as it inherits the standard error of 0.46 of the static IV estimator.

7 Conclusion

In this paper we have introduced a score-driven filter for time-varying regression parameters that can be applied when regressors are endogenous. We have established invertibility of the filter, consistency of the MLE and proven filter convergence to the true unobserved path. In a simulation study we have shown that the behavior of the regular score filter gives undesirable results while the IV-score, in the presence of suitable instruments, uncovers the true underlying path of the time varying parameter. We have also shown in simulations that for non-stationary parameter processes such as structural breaks and random walks the filter shows appropriate behavior. Nevertheless, caution should be taken with extrapolating this result, as non-stationary paths are not included in the theoretical framework.

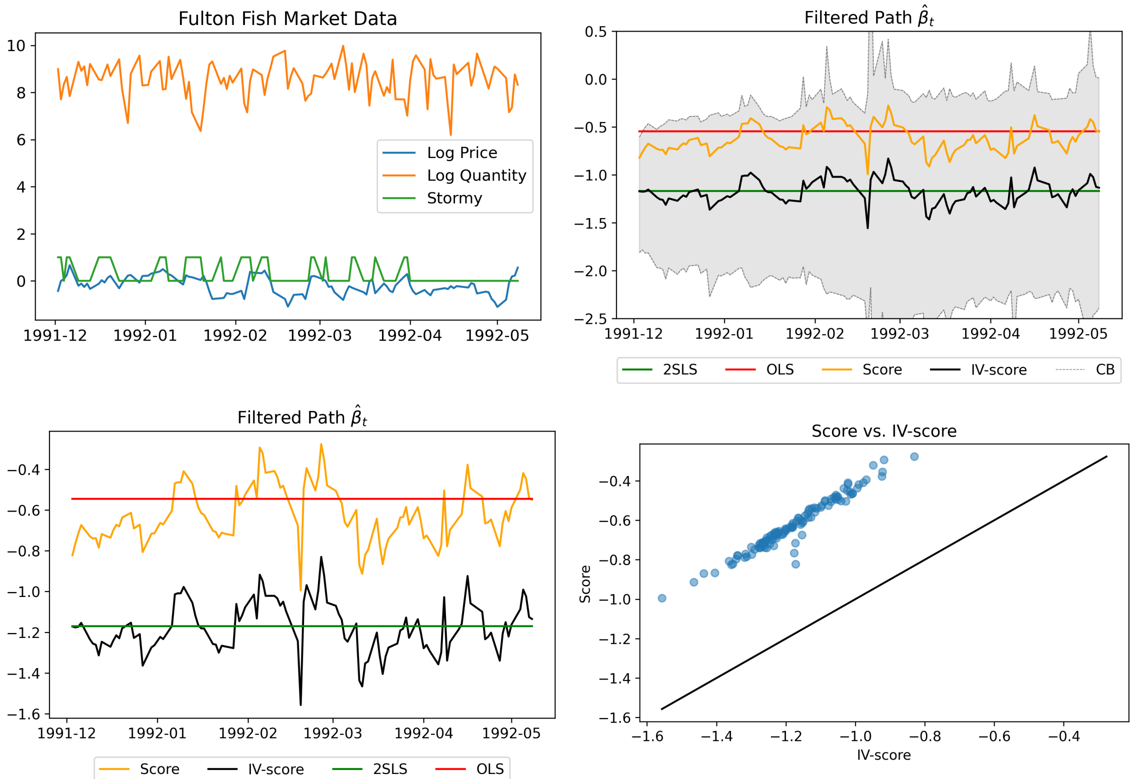

Figure 5: Time-varying price elasticity of demand at Fulton fish market

In an application to excess sensitivity of consumption to income we show that even when static OLS and IV estimates are not significantly different, our analogous time-varying filters could still point out local differences that lead to valuable insights. While the filters coincided for most periods, the few instances in which they did diverge suggest that the level of endogeneity also varies over time, leaving the IV-score filter as the most reliable of the two.

Further research will focus on strengthening the theory for inference for the static parameters and the filtered path. With current assumptions, the derivation of asymptotic normality of the MLE is complicated by the fact that the likelihood function does not have guaranteed bounded moments. With asymptotic normality, a test for time variation and a direct Hausman test for endogeneity could also be developed. The score-driven IV method enriched with such hypothesis tests could significantly add by presenting new results in many fields of applied research.

References

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American economic review*, *93*(1), 113–132.

Angrist, J. D., Graddy, K., & Imbens, G. W. (2000, 07). The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish. *The Review of Economic Studies*, *67*(3), 499-527.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, *91*(434), 444–455.

Ashenfelter, O., & Card, D. (1985). Using the longitudinal structure of earnings to estimate the effect of training programs. *The Review of Economics and Statistics*, *67*(4), 648–660.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, *119*(1), 249–275.

Beutner, E. A., Lin, Y., & Lucas, A. (2023). *Consistency, distributional convergence, and optimality of score-driven filters* (Tech. Rep.). Tinbergen Institute.

Bhatt, V., Kundan Kishor, N., & Marfatia, H. (2020). Estimating excess sensitivity and habit persistence in consumption using greenbook forecasts. *Oxford Bulletin of Economics and Statistics*, *82*(2), 257–284.

Blanchard, O. J., & Leigh, D. (2013). Growth forecast errors and fiscal multipliers. *American Economic Review*, *103*(3), 117–120.

Blasques, F., Francq, C., & Laurent, S. (2022). *Autoregressive conditional betas* [Working paper].

Blasques, F., Francq, C., & Laurent, S. (2023). Quasi score-driven models. *Journal of Econometrics*, *234*(1), 251–275.

Blasques, F., Gorgi, P., Koopman, S. J., & Wintenberger, O. (2018). Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics*, *12*(1), 1019–1052.

Blasques, F., Koopman, S. J., & Lucas, A. (2014). *Maximum likelihood estimation for generalized autoregressive score models* (Tech. Rep.). Tinbergen Institute Discussion Paper.

Blasques, F., Koopman, S. J., & Lucas, A. (2015). Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika*, *102*(2), 325–343.

Blasques, F., Koopman, S. J., & Lucas, A. (2018). Amendments and corrections: 'information-theoretic optimality of observation-driven time series models for continuous responses'. *Biometrika*, *105*(3), 753–753.

Blasques, F., Koopman, S. J., & Lucas, A. (2020). Nonlinear autoregressive models with optimality properties. *Econometric Reviews*, *39*(6), 559–578.

Blasques, F., Lucas, A., & van Vlodrop, A. C. (2021). Finite sample optimality of score-driven volatility models: Some monte carlo evidence. *Econometrics and Statistics*, *19*, 47–57.

Blasques, F., van Brummelen, J., Koopman, S. J., & Lucas, A. (2022). Maximum likelihood estimation for score-driven models. *Journal of Econometrics*, *227*(2), 325–346.

Boivin, J., & Giannoni, M. P. (2006). Has Monetary Policy Become More Effective? *The Review of Economics and Statistics*, *88*(3), 445-462.

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. In *Handbook of econometrics* (Vol. 5, pp. 3705–3843). Elsevier.

Calvori, F., Creal, D., Koopman, S. J., & Lucas, A. (2017). Testing for parameter instability across different modeling frameworks. *Journal of Financial Econometrics*, *15*(2), 223–246.

Campbell, J. Y., & Mankiw, N. G. (1989). Consumption, income, and interest rates: Reinterpreting the time series evidence. *NBER macroeconomics annual*, *4*, 185–216.

Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, *28*(5), 777–795.

Creal, D., Koopman, S. J., Lucas, A., & Zamojski, M. (2018). Generalized autoregressive method of moments.

Eggertsson, G. B., & Krugman, P. (2012). Debt, deleveraging, and the liquidity trap: A fisher-minsky-koo approach. *The Quarterly Journal of Economics*, *127*(3), 1469–1513.

Giraitis, L., Kapetanios, G., & Marcellino, M. (2021). Time-varying instrumental variable estimation. *Journal of Econometrics*, *224*(2), 394–415.

Gorgi, P., Koopman, S. J., & Schaumburg, J. (2017). Time-varying vector autoregressive models with structural dynamic factors. *Tinbergen Institute, The Netherlands and Aarhus University, Denmark*, 17.

Graddy, K. (1995). Testing for imperfect competition at the fulton fish market. *The RAND Journal of Economics*, *26*(1), 75–92.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica, Journal of the Econometric Society*, 1–12.

Hall, R. E. (1978). Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence. *Journal of political economy*, *86*(6), 971–987.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series* (Vol. 52). Cambridge University Press.

Heckman, J. J., & Robb, R. (1985). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics*, *30*(1), 239-267. Retrieved from https://www.sciencedirect.com/science/article/pii/0304407685901393 doi: https://doi.org/10.1016/0304-4076(85)90139-3

Inoue, A., Rossi, B., & Wang, Y. (2022). Local projections in unstable environments: How effective is fiscal policy?

Kennan, J. (1989). Simultaneous equations bias in disaggregated econometric models. *The Review of Economic Studies*, *56*(1), 151–156.

Kim. (2006). Time-varying parameter models with endogenous regressors. *Economics letters*, *91*(1), 21–26.

Kim, & Kim. (2011). Dealing with endogeneity in a time-varying parameter model: joint estimation and two-step estimation procedures. *The Econometrics Journal*, *14*(3), 487–497.

Koop, G., Leon-Gonzalez, R., & Strachan, R. W. (2009). On the evolution of the monetary policy transmission mechanism. *Journal of Economic Dynamics and Control*, *33*(4), 997-1017.

Koopman, S. J., Lucas, A., & Scharth, M. (2016). Predicting time-varying parameters with parameter-driven and observation-driven models. *Review of Economics and Statistics*, *98*(1), 97–110.

Korobilis, D. (2013). Assessing the transmission of monetary policy using time-varying parameter dynamic factor models. *Oxford Bulletin of Economics and Statistics*, *75*(2), 157-179.

Krengel, U. (1985). *Ergodic theorems.* Berlin: De Gruyter.

Müller, U. K., & Petalas, P.-E. (2010). Efficient estimation of the parameter path in unstable time series models. *The Review of Economic Studies*, *77*(4), 1508–1539.

Pascual, L., Romo, J., & Ruiz, E. (2006). Bootstrap prediction for returns and volatilities in garch models. *Computational Statistics Data Analysis*, *50*(9), 2293-2312. (Statistical signal extraction and filtering) doi: https://doi.org/10.1016/j.csda.2004.12.008

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.

Terza, J. V., Basu, A., & Rathouz, P. J. (2008). Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of health economics*, *27*(3), 531–543.

Thiele, S., & Harvey, A. (2013). *Time varying regression and changing correlation.* Retrieved from https://personal.vu.nl/s.j.koopman/projects/ScoreWorkShop2013/Steve%20Thiele%20talk.pdf (Score Workshop Vrije Universiteit Amsterdam)

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational psychology*, *51*(6), 309.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics*, *20*(4), 595–601.

Wooldridge, J. M. (2002). M-estimation. In *Econometric analysis of cross section and panel data.* Cambridge: MIT Press.

Wooldridge, J. M. (2009). Omitted variable bias: the simple case. *Introductory Econometrics: A Modern Approach. Mason, OH: Cengage Learning*, 89–93.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, *50*(2), 420–445.

Wright, P. G. (1928). *The tariff on animal and vegetable oils* (No. 26). New York: Macmillan.

Appendix A: Proofs

Proof of Lemma 1

This lemma is an application of Proposition 3.2 of Blasques, van Brummelen, Koopman, and Lucas (2022), to which we refer for a proof.

Proof of Lemma 2

This lemma is an application of Proposition 3.3 of Blasques, van Brummelen, et al. (2022), to which we refer for a proof.

Proof of Theorem 1

For notational simplicity, write $\hat{\ell}_{T_2}(\boldsymbol{\theta}, \pi, \hat{\beta}_1) := \frac{1}{T_2}\sum_{t=1}^{T_2} \ell_t(\boldsymbol{\theta}, \pi, \hat{\beta}_t(\boldsymbol{\theta}, \pi, \hat{\beta}_1))$ and $\tilde{\ell}_{T_2}(\boldsymbol{\theta}, \pi) := \frac{1}{T_2}\sum_{t=1}^{T_2} \ell_t(\boldsymbol{\theta}, \pi, \tilde{\beta}_t(\boldsymbol{\theta}, \pi))$. We will prove this Theorem in the following parts:

(P1) The model is identifiable: $\ell_\infty(\boldsymbol{\theta}_0, \pi_0) > \ell_\infty(\boldsymbol{\theta}, \pi_0)$ for any $\boldsymbol{\theta} \in \Theta, \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.

(P2) The function $\hat{\ell}_{T_2}(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_1)$ with first stage estimator $\hat{\pi}_{T_1}$ converges in probability to $\tilde{\ell}_{T_2}(\boldsymbol{\theta}, \pi_0)$ uniformly over $\boldsymbol{\Theta}$. That is,

$$\|\hat{\ell}_{T_2}(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_1) - \tilde{\ell}_{T_2}(\boldsymbol{\theta}, \pi_0)\|_\Theta \xrightarrow{p} 0 \text{ as } T_1, T_2 \to \infty.$$

(P3) For any $\varepsilon > 0$, the following inequality holds a.s.

$$\limsup_{T_1, T_2 \to \infty} \sup_{\boldsymbol{\theta} \in B^c(\boldsymbol{\theta}_0, \varepsilon)} \hat{\ell}_{T_2}(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_1) < \ell_\infty(\boldsymbol{\theta}_0, \pi_0),$$

where $B^c(\boldsymbol{\theta}_0, \varepsilon) = \Theta \setminus B(\boldsymbol{\theta}_0, \varepsilon)$ with $B(\boldsymbol{\theta}_0, \varepsilon) = \{\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\theta}_0 - \boldsymbol{\theta}\| < \varepsilon\}$.

(P4) The result in **(P3)** implies consistency.

(P1): Existence of $\ell_\infty(\boldsymbol{\theta}_0, \pi_0)$ is guaranteed by **C2** and by **C6** we have that either $\ell_\infty(\boldsymbol{\theta}, \pi_0) \in \mathbb{R}$ or $\ell_\infty(\boldsymbol{\theta}, \pi_0) = -\infty$. Then

$$\begin{aligned}\ell_0(\theta, \pi_0) - \ell_0(\theta_0, \pi_0) = \log p_\eta &\Big(y_t - \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0)x_t - \tau(x_t - \pi_0 z_t); \lambda\Big) \\ &- \log p_\eta\left(y_t - \beta_t^o x_t - \tau_0(x_t - \pi_0 z_t); \lambda_0\right)\end{aligned}$$

where $\tilde{\beta}_t(\boldsymbol{\theta}, \pi_0)$ is the limit sequence, and $\beta_t^o = \tilde{\beta}_t(\boldsymbol{\theta}_0, \pi_0)$ is the true time-varying parameter due to correct specification of the filter.

$$\begin{aligned}\ell_0(\theta, \pi_0) - \ell_0(\theta_0, \pi_0) &= \log\left(\frac{p_\eta(y_t - \tilde{\beta}_t x_t - \tau(x_t - \pi_0 z_t); \lambda)}{p_\eta(y_t - \beta_t^o x_t - \tau_0(x_t - \pi_0 z_t); \lambda_0)}\right) \\ &\leq \frac{p_\eta\left(y_t - \tilde{\beta}_t x_t - \tau(x_t - \pi_0 z_t); \lambda\right)}{p_\eta(y_t - \beta_t^o x_t - \tau_0(x_t - \pi_0 z_t); \lambda_0)} - 1\end{aligned}$$

The case of an equal sign here is ruled out, since the densities are not the same for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. And, since $p_\eta(y_t - \beta_t^o x_t - \tau_0(x_t - \pi_0 z_t); \lambda_0)$ is the true conditional density we have

$$\mathbb{E}[\mathbb{E}\left[\ell_0(\boldsymbol{\theta}, \pi_0) - \ell_0(\boldsymbol{\theta}_0, \pi_0)|Y^t\right]] < \mathbb{E}\left[\mathbb{E}\left[\frac{p_\eta\left(y_t - \tilde{\beta}_t x_t - \tau(x_t - \pi_0 z_t); \lambda\right)}{p_\eta(y_t - \beta_t^o x_t - \tau_0(x_t - \pi_0 z_t); \lambda_0)}\middle| Y^t\right]\right] - 1 = 0$$

As a result,

$$\ell_\infty(\boldsymbol{\theta}, \pi_0) - \ell_\infty(\boldsymbol{\theta}_0, \pi_0) = \mathbb{E}[\mathbb{E}\left[\ell_0(\boldsymbol{\theta}, \pi_0) - \ell_0(\boldsymbol{\theta}_0, \pi_0)|Y^t\right]] < 0 \quad \forall \boldsymbol{\theta} \neq \boldsymbol{\theta}_0.$$

(P2) $\|\hat{\ell}_{T_2}(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_1) - \tilde{\ell}_{T_2}(\boldsymbol{\theta}, \pi_0)\|_\Theta \xrightarrow{p} 0$ as $T_1, T_2 \to \infty$.
Since π might affect the likelihood function directly as well as indirectly through the filter, we denote this explicitly by allowing for different values of π in $\ell_t(\boldsymbol{\theta}, \pi^*, \beta_t(\boldsymbol{\theta}, \pi^{**}))$, to isolate each effect. We have

$$\begin{aligned}&\|\hat{\ell}_{T_2}(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_1) - \tilde{\ell}_{T_2}(\boldsymbol{\theta}, \pi_0)\|_\Theta \\ &\qquad = \sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{T_2}\sum_{t=1}\ell_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_1)) - \frac{1}{T_2}\sum_{t=1}\ell_t(\boldsymbol{\theta}, \pi_0, \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0))\right| \\ &\qquad \leq \sup_{\boldsymbol{\theta}\in\Theta}\frac{1}{T_2}\sum_{t=1}\left|\ell_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_1)) - \ell_t(\boldsymbol{\theta}, \pi_0, \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0))\right| \\ &\qquad \leq \frac{1}{T_2}\sum_{t=1}\left\|\ell_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_1)) - \ell_t(\boldsymbol{\theta}, \pi_0, \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0))\right\|_\Theta.\end{aligned}$$

We will show that each summand converges in probability to zero, so that by the Stolz-Cesaro theorem the sample average does too. We can write

$$\begin{aligned}&\|\ell_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_1)) - \ell_t(\boldsymbol{\theta}, \pi_0, \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0))\|_\Theta \\ &\qquad \leq \|\ell_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_1)) - \ell_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \tilde{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}))\|_\Theta \\ &\qquad + \|\ell_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \tilde{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1})) - \ell_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0))\|_\Theta \\ &\qquad + \|\ell_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0) - \ell_t(\boldsymbol{\theta}, \pi_0, \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0))\|_\Theta\end{aligned}$$

Applying a Mean Value Theorem (MVT) to each of these terms and further decomposing it we get by (C4) and (C7)

$$\leq \left\| \frac{\partial \ell_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \beta)}{\partial \beta}\bigg|_{\beta=\beta_t^*(\boldsymbol{\theta},\hat{\pi}_{T_1},\hat{\beta}_1)} \right\|_\Theta \underbrace{\|\hat{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_1) - \tilde{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1})\|_\Theta}_{\xrightarrow{e.a.s.} 0 \text{ as } t\to\infty}$$
$$+ \left\| \frac{\partial \ell_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \beta)}{\partial \beta}\bigg|_{\beta=\tilde{\beta}_t^*(\boldsymbol{\theta},\pi)} \right\|_\Theta \left\| \frac{\partial \tilde{\beta}_t(\boldsymbol{\theta}, \pi)}{\partial \pi}\bigg|_{\pi=\pi_{T_1}^*} \right\|_\Theta \underbrace{|\hat{\pi}_{T_1} - \pi_0|}_{\xrightarrow{p} 0 \text{ as } T_1\to\infty}$$
$$+ \left\| \frac{\partial \ell_t(\boldsymbol{\theta}, \pi, \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0))}{\partial \pi}\bigg|_{\pi=\pi_{T_1}^*} \right\|_\Theta \underbrace{|\hat{\pi}_{T_1} - \pi_0|}_{\xrightarrow{p} 0 \text{ as } T_1\to\infty}$$

where

- $\beta_t^*(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_1)$ is a point between $\hat{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_1)$ and $\tilde{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1})$
- $\tilde{\beta}_t^*(\boldsymbol{\theta}, \pi)$ is a point between $\tilde{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1})$ and $\tilde{\beta}_t(\boldsymbol{\theta}, \pi_0)$
- $\pi_{T_1}^*$ is a point between $\hat{\pi}_{T_1}$ and π_0.

The derivatives in lines 2 and 3 are bounded in probability, as (C5) assumes these to be SE and the derivative process of stationary $\tilde{\beta}_t$ is SE as shown in the technical appendix of Blasques, Koopman, and Lucas (2014), as it yields the same contraction condition. Furthermore, the derivative in line 1 is asymptotically bounded in probability, since $\|\beta_t^*(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_1) - \tilde{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1})\|_\Theta \xrightarrow{e.a.s.} 0$, meaning there exists a value $N \in \mathbb{N}$, such that for all $t > N$ $\mathbb{P}(\|\beta_t^*(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_1) - \tilde{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1})\|_\Theta < 1) = 1$. Therefore,

$$\left\| \frac{\partial \ell_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \beta)}{\partial \beta}\bigg|_{\beta=\beta_t^*(\boldsymbol{\theta},\hat{\pi}_{T_1},\hat{\beta}_1)} \right\|_\Theta \leq \sup_{\boldsymbol{\theta},\, \delta\in[-1,1]} \left| \frac{\partial \ell_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \beta)}{\partial \beta}\bigg|_{\beta=\tilde{\beta}_t(\boldsymbol{\theta},\hat{\pi}_{T_1})+\delta} \right|, \quad \forall\, t > N.$$

Since the latter is SE, by continuity of the supremum operator and by Proposition 4.3 in Krengel (1985), we have that the derivative is asympotically bounded in probability, i.e.

$$\lim_{t,M\to\infty} \mathbb{P}\left(\left\| \frac{\partial \ell_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \beta)}{\partial \beta}\bigg|_{\beta=\beta_t^*(\boldsymbol{\theta},\hat{\pi}_{T_1},\hat{\beta}_1)} \right\|_\Theta > M \right)$$
$$\leq \lim_{t,M\to\infty} \mathbb{P}\left(\sup_{\boldsymbol{\theta},\, \delta\in[-1,1]} \left| \frac{\partial \ell_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \beta)}{\partial \beta}\bigg|_{\beta=\tilde{\beta}_t(\boldsymbol{\theta},\hat{\pi}_{T_1})+\delta} \right| > M \right) = 0.$$

(P3): We will show that (**P3**) holds for $\tilde{\ell}_{T_2}(\boldsymbol{\theta}, \pi_0)$ since by (**P2**) $\hat{\ell}_{T_2}(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_1)$ is asymptotically equivalent to $\tilde{\ell}_{T_2}(\boldsymbol{\theta}, \pi_0)$.

Fix a $\boldsymbol{\theta}^*$. Then for a decreasing sequence $\{\varepsilon_i\}_{i\in\mathbb{N}}$ s.t. $\lim_{i\to\infty}\varepsilon_i = 0$, the sequence $\{\sup_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}^*,\varepsilon_i)} \ell_0(\boldsymbol{\theta},\pi_0)\}_{i\in\mathbb{N}}$ is non-increasing and greater than $\ell_0(\boldsymbol{\theta}^*,\pi_0)$ for every i. Considering this, and the fact that $\lim_{i\to\infty}\sup_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}^*,\varepsilon_i)} \ell_0(\boldsymbol{\theta},\pi_0) = \ell_0(\boldsymbol{\theta}^*,\pi_0)$ by continuity, we conclude that $\sup_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}^*,\varepsilon_i)} \ell_0(\boldsymbol{\theta},\pi_0) \downarrow \ell_0(\boldsymbol{\theta}^*,\pi_0)$. This together with $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\ell_0(\boldsymbol{\theta},\pi_0) < \infty$ which is implied by (**C6**), we can apply the Monotone Convergence Theorem to conclude that

$$\lim_{i\to\infty}\mathbb{E}\sup_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}^*,\varepsilon_i)} \ell_0(\boldsymbol{\theta},\pi_0) = \ell_\infty(\boldsymbol{\theta}^*,\pi_0).$$

By (**P1**) we have that $\ell_\infty(\boldsymbol{\theta}_0,\pi_0) > \ell_\infty(\boldsymbol{\theta}^*,\pi_0)$ so that for all $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$ there exists a $\varepsilon_{\boldsymbol{\theta}^*} > 0$ such that

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}^*,\varepsilon_{\boldsymbol{\theta}^*})} \ell_0(\boldsymbol{\theta},\pi_0) < \ell_\infty(\boldsymbol{\theta}_0,\pi_0).$$

The set $B^c(\boldsymbol{\theta}_0,\varepsilon)$ is compact and is covered by the balls $\{B(\boldsymbol{\theta},\varepsilon_{\boldsymbol{\theta}}) : \boldsymbol{\theta}\in B^c(\boldsymbol{\theta}_0,\varepsilon)\}$. Let $B(\boldsymbol{\theta}_1,\varepsilon_1),\ldots,B(\boldsymbol{\theta}_p,\varepsilon_p)$ be a finite subcover with $\sup_{k=1,\ldots,p}\varepsilon_k < \varepsilon$. Then, for any $T\in\mathbb{N}$, we have

$$\sup_{\boldsymbol{\theta}\in B^c(\boldsymbol{\theta}_0,\varepsilon)} \tilde{\ell}_T(\boldsymbol{\theta},\pi_0) \le \bigvee_{k=1}^{p}\frac{1}{T}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}_k,\varepsilon_k)} \tilde{\ell}_t(\boldsymbol{\theta},\pi_0).$$

Taking limits on both sides of the equation gives

$$\limsup_{T_2\to\infty}\sup_{\boldsymbol{\theta}\in B^c(\boldsymbol{\theta}_0,\varepsilon)} \tilde{\ell}_{T_2}(\boldsymbol{\theta},\pi_0) \le \bigvee_{k=1}^{p}\mathbb{E}\sup_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}_k,\varepsilon_k)} \ell_0(\boldsymbol{\theta},\pi_0) < \ell_\infty(\boldsymbol{\theta}_0,\pi_0).$$

(P4): If there exists an $\varepsilon > 0$ such that $|\hat{\boldsymbol{\theta}}_{T_2}(\hat{\pi}_{T_1},\hat{\beta}_1) - \boldsymbol{\theta}_0| > \varepsilon$, that implies

$$\sup_{\boldsymbol{\theta}\in B^c_{\varepsilon_{\boldsymbol{\theta}_0}}} \hat{\ell}_{T_2}(\boldsymbol{\theta},\hat{\pi}_{T_1},\hat{\beta}_1) \ge \hat{\ell}_{T_2}\left(\hat{\boldsymbol{\theta}}_{T_2}(\hat{\pi}_{T_1},\hat{\beta}_1),\hat{\pi}_{T_1},\hat{\beta}_1\right) \ge \hat{\ell}_{T_2}(\boldsymbol{\theta}_0,\hat{\pi}_{T_1},\hat{\beta}_1)$$

by definition of the two-step MLE. Therefore,

$$\lim_{T_1,T_2\to\infty}\mathbb{P}\left(|\hat{\boldsymbol{\theta}}_{T_2}(\hat{\pi}_{T_1},\hat{\beta}_1) - \boldsymbol{\theta}_0| > \varepsilon\right) \le \lim_{T_1,T_2\to\infty}\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in B^c_{\varepsilon_{\boldsymbol{\theta}_0}}} \hat{\ell}_{T_2}(\boldsymbol{\theta},\hat{\pi}_{T_1},\hat{\beta}_1) - \hat{\ell}_{T_2}(\boldsymbol{\theta}_0,\hat{\pi}_{T_1},\hat{\beta}_1) \ge 0\right).$$

Furthermore,

$$\begin{aligned}
0 &\le \sup_{\boldsymbol{\theta}\in B^c_{\varepsilon_{\boldsymbol{\theta}_0}}} \hat{\ell}_{T_2}(\boldsymbol{\theta},\hat{\pi}_{T_1},\hat{\beta}_1) - \hat{\ell}_{T_2}(\boldsymbol{\theta}_0,\hat{\pi}_{T_1},\hat{\beta}_1) \\
&\le \sup_{\boldsymbol{\theta}\in B^c_{\varepsilon_{\boldsymbol{\theta}_0}}} \hat{\ell}_{T_2}(\boldsymbol{\theta},\hat{\pi}_{T_1},\hat{\beta}_1) - \tilde{\ell}_{T_2}(\boldsymbol{\theta}_0,\pi_0) + \tilde{\ell}_{T_2}(\boldsymbol{\theta}_0,\pi_0) - \hat{\ell}_{T_2}(\boldsymbol{\theta}_0,\hat{\pi}_{T_1},\hat{\beta}_1) \\
&\le \sup_{\boldsymbol{\theta}\in B^c_{\varepsilon_{\boldsymbol{\theta}_0}}} \hat{\ell}_{T_2}(\boldsymbol{\theta},\hat{\pi}_{T_1},\hat{\beta}_1) - \tilde{\ell}_{T_2}(\boldsymbol{\theta}_0,\pi_0) + |\tilde{\ell}_{T_2}(\boldsymbol{\theta}_0,\pi_0) - \hat{\ell}_{T_2}(\boldsymbol{\theta}_0,\hat{\pi}_{T_1},\hat{\beta}_1)| \\
&\le \sup_{\boldsymbol{\theta}\in B^c_{\varepsilon_{\boldsymbol{\theta}_0}}} \hat{\ell}_{T_2}(\boldsymbol{\theta},\hat{\pi}_{T_1},\hat{\beta}_1) - \tilde{\ell}_{T_2}(\boldsymbol{\theta}_0,\pi_0) + \|\tilde{\ell}_{T_2}(\boldsymbol{\theta},\pi_0) - \hat{\ell}_{T_2}(\boldsymbol{\theta},\hat{\pi}_{T_1},\hat{\beta}_1)\|_\Theta \\
&= \sup_{\boldsymbol{\theta}\in B^c_{\varepsilon_{\boldsymbol{\theta}_0}}} \hat{\ell}_{T_2}(\boldsymbol{\theta},\hat{\pi}_{T_1},\hat{\beta}_1) - \tilde{\ell}_{T_2}(\boldsymbol{\theta}_0,\pi_0) + o_p(1)
\end{aligned}$$

by **(P2)**. Then,

$$\begin{aligned}
&\lim_{T_1,T_2\to\infty} \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in B^c_{\varepsilon_{\boldsymbol{\theta}_0}}} \hat{\ell}_{T_2}(\boldsymbol{\theta},\hat{\pi}_{T_1},\hat{\beta}_1) - \hat{\ell}_{T_2}(\boldsymbol{\theta}_0,\hat{\pi}_{T_1},\hat{\beta}_1) \ge 0\right) \\
&\le \lim_{T_1,T_2\to\infty} \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in B^c_{\varepsilon_{\boldsymbol{\theta}_0}}} \hat{\ell}_{T_2}(\boldsymbol{\theta},\hat{\pi}_{T_1},\hat{\beta}_1) - \tilde{\ell}_{T_2}(\boldsymbol{\theta}_0,\pi_0) + o_p(1) \ge 0\right) \quad \text{(ineq above)} \\
&\le \limsup_{T_1,T_2\to\infty} \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in B^c_{\varepsilon_{\boldsymbol{\theta}_0}}} \hat{\ell}_{T_2}(\boldsymbol{\theta},\hat{\pi}_{T_1},\hat{\beta}_1) - \tilde{\ell}_{T_2}(\boldsymbol{\theta}_0,\pi_0) + o_p(1) \ge 0\right) \quad (\lim \le \text{limsup}) \\
&\le \mathbb{P}\left(\limsup_{T_1,T_2\to\infty}\left[\sup_{\boldsymbol{\theta}\in B^c_{\varepsilon_{\boldsymbol{\theta}_0}}} \hat{\ell}_{T_2}(\boldsymbol{\theta},\hat{\pi}_{T_1},\hat{\beta}_1) - \tilde{\ell}_{T_2}(\boldsymbol{\theta}_0,\pi_0) + o_p(1)\right] \ge 0\right) \quad \text{(reverse Fatou's lemma)} \\
&\le \mathbb{P}\left(\limsup_{T_1,T_2\to\infty}\sup_{\boldsymbol{\theta}\in B^c_{\varepsilon_{\boldsymbol{\theta}_0}}} \hat{\ell}_{T_2}(\boldsymbol{\theta},\hat{\pi}_{T_1},\hat{\beta}_1) \ge \limsup_{T_1,T_2\to\infty} \tilde{\ell}_{T_2}(\boldsymbol{\theta}_0,\pi_0) - \limsup_{T_1,T_2\to\infty} o_p(1)\right) \quad \text{(see note)} \\
&= \mathbb{P}\left(\limsup_{T_1,T_2\to\infty}\sup_{\boldsymbol{\theta}\in B^c_{\varepsilon_{\boldsymbol{\theta}_0}}} \hat{\ell}_{T_2}(\boldsymbol{\theta},\hat{\pi}_{T_1},\hat{\beta}_1) \ge \ell_\infty(\boldsymbol{\theta}_0,\pi_0)\right) = 0 \quad \textbf{(P3)}
\end{aligned}$$

Note: The rule $\limsup_{T\to\infty}(a_T+b_T) \le \limsup_{T\to\infty} a_T + \limsup_{T\to\infty} b_T$ holds for bounded sequences, which is guaranteed by **(P3)** and (C2).

Proof of Corollary 3

In order to show consistency of the MLE for the Gaussian IV-score model, we show that the assumptions imply each of the conditions *C1-C7* from Theorem 1. Let $K = \frac{1}{2}\log(2\Pi)$

where Π is used here to denote the number pi (3.14..) that appears in the Gaussian density. Then we have the following expressions for the likelihood contributions and the IV-score filter

$$\ell_t(\boldsymbol{\theta},\pi,\hat{\beta}_t(\boldsymbol{\theta},\pi,\hat{\beta}_1)) = -K - \frac{1}{2}\log(\sigma^2) - \frac{1}{2}\sigma^{-2}(y_t - \hat{\beta}_t(\boldsymbol{\theta},\pi,\hat{\beta}_1)x_t - \tau(x_t - \pi z_t))^2$$
$$\hat{\beta}_{t+1}(\boldsymbol{\theta},\pi,\hat{\beta}_1) = \omega + \alpha\sigma^{-2}x_t(y_t - \hat{\beta}_t(\boldsymbol{\theta},\pi,\hat{\beta}_1)x_t - \tau(x_t - \pi z_t)) + \gamma\hat{\beta}_t(\boldsymbol{\theta},\pi,\hat{\beta}_1),$$

where $\hat{\beta}_1(\boldsymbol{\theta},\pi,\hat{\beta}_1) = \hat{\beta}_1$. Furthermore, we define

$$\ell_0(\boldsymbol{\theta},\pi) = -K - \frac{1}{2}\log(\sigma^2) - \frac{1}{2}\sigma^{-2}(y_t - \tilde{\beta}_t(\boldsymbol{\theta},\pi)x_t - \tau(x_t - \pi z_t))^2, \tag{16}$$

where $\tilde{\beta}_t(\boldsymbol{\theta},\pi)$ is the stationary limit of $\hat{\beta}_t(\boldsymbol{\theta},\pi,\hat{\beta}_1)$.

(C1) The DGP admits a stationary solution if $\beta_t^o, x_t, z_t, \eta_t$ are stationary sequences by Proposition 4.3 in Krengel (1985). Therefore we need to show that $\beta_t^0 = \tilde{\beta}_t(\boldsymbol{\theta}_0,\pi_0)$ admits a stationary solution, since the rest is assumed to be stationary. The true parameter follows the process $\beta_{t+1}^0 = \omega_0 + \alpha_0\sigma_0^2 x_t\eta_t + \gamma_0\beta_t^0$ which is SE whenever $E|\gamma_0| < 1$.

(C2) $\mathbb{E}|\ell_0(\boldsymbol{\theta}_0,\pi_0)| < \infty$ holds for the Gaussian density.

$$\begin{aligned}\mathbb{E}|\ell_0(\boldsymbol{\theta}_0,\pi_0)| &= \mathbb{E}|-K - \frac{1}{2}\log(\sigma_0^2) - \frac{1}{2}\sigma_0^{-2}(y_t - \beta_t^0 x_t - \tau_0 u_t)^2| \\ &= \mathbb{E}|-K - \frac{1}{2}\log(\sigma_0^2) - \frac{1}{2}\sigma_0^{-2}\eta_t^2| < \infty.\end{aligned}$$

(C3) We have to show that $\ell_0(\boldsymbol{\theta}_0,\pi_0) = \ell_0(\boldsymbol{\theta},\pi_0)$ if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Since the "if" direction is obvious, we focus on the "only if" component. Denote the Gaussian density function by $f(y|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{1}{2}(\frac{y-\mu}{\sigma})^2)$. Note that $f(y|\mu_0,\sigma_0) = f(y|\mu,\sigma)$ for any y if and only if $\mu = \mu_0$ and $\sigma = \sigma_0$. Therefore, for our Gaussian likelihood in equation (16) with $\sigma = \sigma_0$, we have $\mu = \mu_0$ if and only if

$$\tilde{\beta}_t(\boldsymbol{\theta}_0,\pi_0)x_t + \tau_0(x_t - \pi_0 z_t) = \tilde{\beta}_t(\boldsymbol{\theta},\pi_0)x_t + \tau(x_t - \pi_0 z_t).$$

Using the first stage equation $x_t = \pi_0 z_t + u_t$ we can rearrange this to

$$0 = \pi_0(\tilde{\beta}_t(\boldsymbol{\theta}_0,\pi_0) - \tilde{\beta}_t(\boldsymbol{\theta},\pi_0))z_t + (\tilde{\beta}_t(\boldsymbol{\theta}_0,\pi_0) - \tilde{\beta}_t(\boldsymbol{\theta},\pi_0) + (\tau_0 - \tau))u_t.$$

As $u_t \perp z_t$ this can only hold for any z_t, u_t if the terms premultiplying z_t and u_t are both zero. Solving the system of these two equations we get

$$\pi_0(\tau_0 - \tau) = 0,$$

with solutions *i)* $\tau = \tau_0$ and *ii)* $\pi_0 = 0$ and possibly $\tau \neq \tau_0$ (which means also $\tilde{\beta}_t(\boldsymbol{\theta}_0,\pi_0) - \tilde{\beta}_t(\boldsymbol{\theta},\pi_0) = \tau - \tau_0 \neq 0$). It remains to verify what these solutions imply for the rest of the elements in $\boldsymbol{\theta}$. To investigate both solutions simultaneously, define $\kappa := \tilde{\beta}_t(\boldsymbol{\theta}_0, \pi_0) - \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0)$, so that for solution $i)$ $\kappa = 0$ and $ii)$ $\kappa = \tau - \tau_0 \neq 0$ for any t. Then for the filter at time $t+1$,

$$
\begin{aligned}
\kappa &= \tilde{\beta}_{t+1}(\boldsymbol{\theta}_0, \pi_0) - \tilde{\beta}_{t+1}(\boldsymbol{\theta}, \pi_0) \\
\kappa &= (\omega_0 + \alpha_0 \sigma_0^{-2} x_t \eta_t + \gamma_0 \tilde{\beta}_t(\boldsymbol{\theta}_0, \pi_0) \\
&\quad -(\omega + \alpha \sigma_0^{-2} x_t \underbrace{(y_t - \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0) x_t - \tau(x_t - \pi_0 z_t))}_{\tilde{\eta}_t} + \gamma \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0) \\
\kappa &= (\omega_0 - \omega) + \frac{(\alpha_0 \eta_t - \alpha \tilde{\eta}_t)}{\sigma_0^2} x_t + (\gamma_0 \tilde{\beta}_t(\boldsymbol{\theta}_0, \pi_0) - \gamma \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0)) \\
0 &= (\omega_0 - \omega) + \frac{(\alpha_0 \eta_t - \alpha \tilde{\eta}_t)}{\sigma_0^2} x_t + (\gamma_0 - \gamma) \tilde{\beta}_t(\boldsymbol{\theta}_0, \pi_0) + (\gamma - 1)\kappa.
\end{aligned}
$$

Since the filter depends on the past, x_t and $\tilde{\beta}_t(\boldsymbol{\theta}_0, \pi_0)$ are independent, meaning that all premultiplying terms should be zero for this to hold. This gives rise to the following set of equations:

$$
\begin{aligned}
(\omega_0 - \omega) + (\gamma - 1)\kappa &= 0 \\
\alpha_0 \eta_t - \alpha \tilde{\eta}_t &= 0 \\
\gamma_0 - \gamma &= 0.
\end{aligned}
$$

Since we have the solution $\pi_0(\tau_0 - \tau) = 0$, the second equation gives us

$$
\begin{aligned}
0 = \alpha_0 \eta_t - \alpha \tilde{\eta}_t &= \alpha_0 \eta_t - \alpha(y_t - \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0) x_t - \tau(x_t - \pi_0 z_t)) \\
&= (\alpha_0 - \alpha)\eta_t - \alpha \left[\left(\tilde{\beta}_t(\boldsymbol{\theta}_0, \pi_0) - \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0) + \tau_0 - \tau \right) x_t - (\tau_0 \pi_0 - \tau \pi_0) z_t \right] \\
&= (\alpha_0 - \alpha)\eta_t - \alpha \left[\left(\tilde{\beta}_t(\boldsymbol{\theta}_0, \pi_0) - \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0) + \tau_0 - \tau \right) \pi_0 - (\tau_0 \pi_0 - \tau \pi_0) \right] z_t \\
&- \alpha \left(\tilde{\beta}_t(\boldsymbol{\theta}_0, \pi_0) - \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0) + \tau_0 - \tau \right) u_t \\
&= (\alpha_0 - \alpha)\eta_t \implies \alpha_0 - \alpha = 0.
\end{aligned}
$$

Then we have (including the previously established values)

$$
\begin{aligned}
\sigma &= \sigma_0 \\
\gamma &= \gamma_0 \\
\alpha &= \alpha_0 \\
(\omega_0 - \omega) + (\gamma_0 - 1)(\tau - \tau_0) &= 0 \\
\pi_0(\tau - \tau_0) &= 0.
\end{aligned}
$$

Since $\pi_0 \neq 0$, we finally obtain that $\tau = \tau_0, \omega = \omega_0$ meaning that $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

(C4) This is a direct result of Corollary 1.

(C5) Writing out each derivative we obtain

$$\frac{\partial \ell_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \tilde{\beta}_t^*(\boldsymbol{\theta}, \pi))}{\partial \beta} = \frac{x_t(y_t - \tilde{\beta}_t^*(\boldsymbol{\theta}, \pi)x_t - \tau(x_t - \hat{\pi}_{T_1} z_t))}{\sigma^2}$$
$$\frac{\partial \ell_t(\boldsymbol{\theta}, \pi_{T_1}^*, \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0))}{\partial \pi} = -\frac{\tau z_t(y_t - \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0)x_t - \tau(x_t - \pi_{T_1}^* z_t))}{\sigma^2},$$

which are continuous functions of SE sequences.

(C6) Implied by Gaussian density and boundedness of the parameter space.

$$\begin{aligned}\ell_0(\boldsymbol{\theta}, \pi_0) &= -K - \frac{1}{2}\log(\sigma) - \frac{1}{2}\sigma^{-2}(y_t - \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0)x_t - \tau(x_t - \pi_0 z_t))^2 \\ &\leq -K - \frac{1}{2}\log(\sigma) \qquad \text{a.s.}\end{aligned}$$

Then

$$\sup_{\boldsymbol{\theta}\in\Theta} \ell_0(\boldsymbol{\theta}, \pi_0) \leq -K - \frac{1}{2}\sup_{\boldsymbol{\theta}\in\Theta} \log \sigma < \infty,$$

with probability 1 by the compactness of Θ. This also implies $E\|\ell_0(\boldsymbol{\theta}, \pi_0) \vee c\|_\Theta < \infty$.

(C7) Given the DGP with Gaussian i.i.d. errors, the first stage OLS estimator is consistent.

Proof of Proposition 1

We assume correct specification, which implies that $\beta_t^o = \tilde{\beta}_t(\boldsymbol{\theta}_0, \pi_0)$. To make dependencies explicit, denote the whole updating equation by $\beta_{t+1}(\boldsymbol{\theta}, \pi) = \phi(\beta_t(\boldsymbol{\theta}, \pi), Y_t, \pi; \boldsymbol{\theta}))$. Finally, let $B_\varepsilon(\boldsymbol{\theta}) = \{\tilde{\boldsymbol{\theta}} \in \Theta : \|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\| \leq \varepsilon\}$ be a compact neighborhood of $\boldsymbol{\theta} \in \Theta$ with radius ε for some $\varepsilon > 0$. Then,

$$\begin{aligned}&|\hat{\beta}_t(\hat{\boldsymbol{\theta}}_{T_2}, \hat{\pi}_{T_1}, \hat{\beta}_1) - \beta_t^o| = |\hat{\beta}_t(\hat{\boldsymbol{\theta}}_{T_2}, \hat{\pi}_{T_1}, \hat{\beta}_1) - \tilde{\beta}_t(\boldsymbol{\theta}_0, \pi_0)| \\ &\leq |\hat{\beta}_t(\hat{\boldsymbol{\theta}}_{T_2}, \hat{\pi}_{T_1}, \hat{\beta}_1) - \tilde{\beta}_t(\hat{\boldsymbol{\theta}}_{T_2}, \hat{\pi}_{T_1})| + |\tilde{\beta}_t(\hat{\boldsymbol{\theta}}_{T_2}, \hat{\pi}_{T_1}) - \tilde{\beta}_t(\hat{\boldsymbol{\theta}}_{T_2}, \pi_0)| + |\tilde{\beta}_t(\hat{\boldsymbol{\theta}}_{T_2}, \pi_0) - \tilde{\beta}_t(\boldsymbol{\theta}_0, \pi_0)| \\ &\leq \|\hat{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}, \hat{\beta}_1) - \tilde{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1})\|_\Theta + \|\tilde{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}) - \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0)\|_\Theta + |\tilde{\beta}_t(\hat{\boldsymbol{\theta}}_{T_2}, \pi_0) - \tilde{\beta}_t(\boldsymbol{\theta}_0, \pi_0)|\end{aligned}$$

The first term vanishes e.a.s. as $t \to \infty$ by an application of Lemma 1. The second term can be shown to vanish a.s. by an application of the MVT, similar to the proof of **(P2)** in Theorem 1.

$$\|\tilde{\beta}_t(\boldsymbol{\theta}, \hat{\pi}_{T_1}) - \tilde{\beta}_t(\boldsymbol{\theta}, \pi_0)\|_\Theta \leq \underbrace{\left\|\frac{\partial \tilde{\beta}_t(\boldsymbol{\theta}, \pi_{T_1}^*)}{\partial \pi}\right\|_\Theta}_{SE} \underbrace{|\hat{\pi}_{T_1} - \pi_0|}_{\xrightarrow{p} 0}$$

where $\pi^*_{T_1}$ is a point between $\hat{\pi}_{T_1}$ and π_0, and the derivative process of stationary $\tilde{\beta}_t$ is SE (hence, bounded in probability) as shown in the technical appendix of Blasques et al. (2014). A similar argument holds for the third term, where we expand around $\boldsymbol{\theta}^*_{T_2}$, a point in between $\hat{\boldsymbol{\theta}}_{T_2}$ and $\boldsymbol{\theta}_0$.

$$|\tilde{\beta}_t(\hat{\boldsymbol{\theta}}_{T_2}, \pi_0) - \tilde{\beta}_t(\boldsymbol{\theta}_0, \pi_0)| \leq \underbrace{\left\| \frac{\partial \tilde{\beta}_t(\boldsymbol{\theta}^*_{T_2}, \pi)}{\partial \boldsymbol{\theta}} \right\|_\Theta}_{SE} \underbrace{|\hat{\boldsymbol{\theta}}_{T_2} - \boldsymbol{\theta}_0|}_{\xrightarrow{p} 0} \xrightarrow{p} 0.$$

Appendix B: Additional Simulation Results

Figure 6: β_t with a midway break (DGP 2).

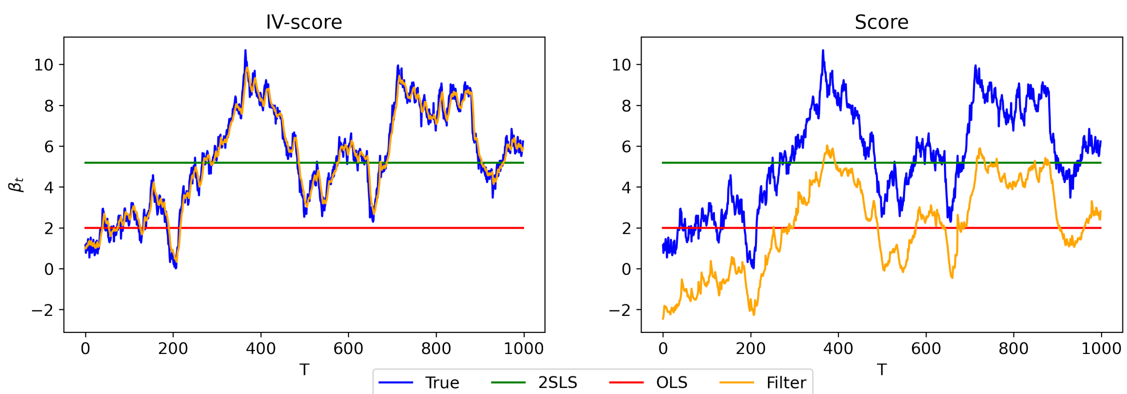

Figure 7: β_t following a random walk (DGP 3).

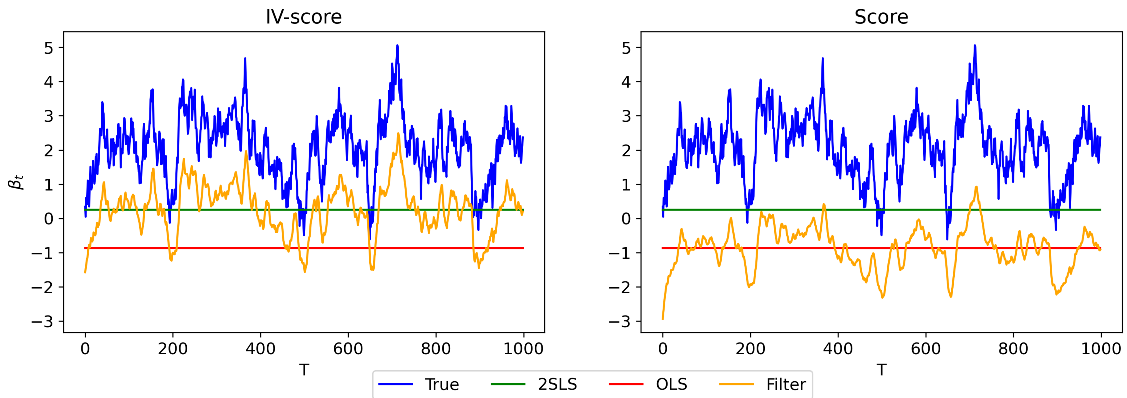

Figure 8: Filtering with an endogenous instrument (DGP 4).

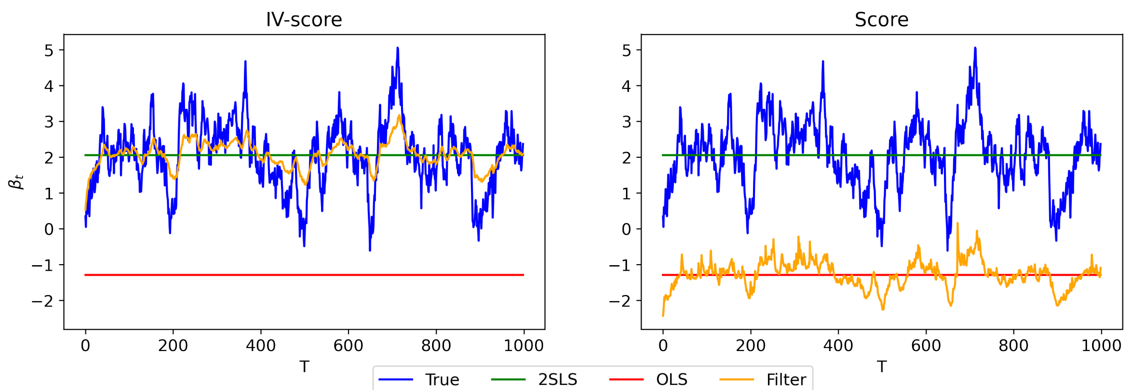

Figure 9: Filtering with large error variance (DGP 5).

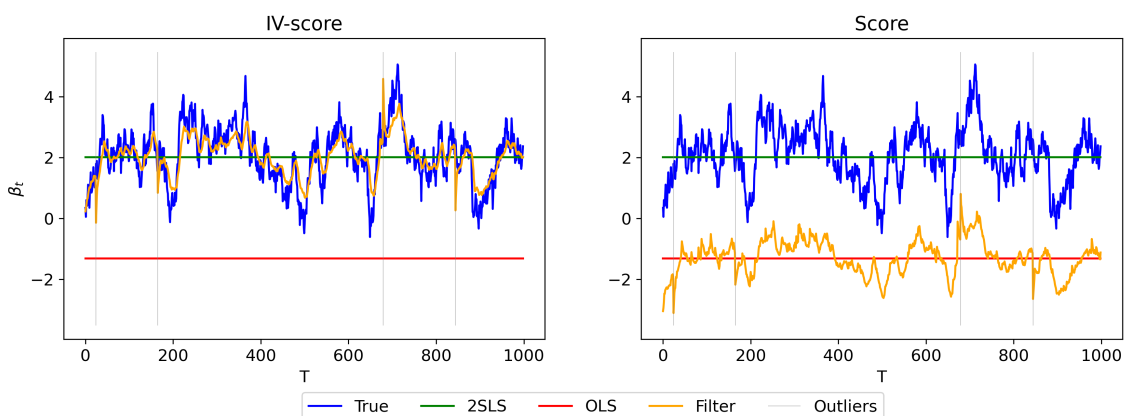

Figure 10: Filtering with outliers (DGP 6).