High-Dimensional Mixed-Frequency IV Regression

Andrii Babii
Department of Economics, University of North Carolina at Chapel Hill, Chapel Hill, NC

ABSTRACT
This article introduces a high-dimensional linear IV regression for the data sampled at mixed frequencies. We show that the high-dimensional slope parameter of a high-frequency covariate can be identified and accurately estimated leveraging on a low-frequency instrumental variable. The distinguishing feature of the model is that it allows handling high-dimensional datasets without imposing the approximate sparsity restrictions. We propose a Tikhonov-regularized estimator and study its large sample properties for time series data. The estimator has a closed-form expression that is easy to compute and demonstrates excellent performance in our Monte Carlo experiments. We also provide the confidence bands and incorporate the exogenous covariates via the double/debiased machine learning approach. In our empirical illustration, we estimate the real-time price elasticity of supply on the Australian electricity spot market. Our estimates suggest that the supply is relatively inelastic throughout the day.

1. Introduction
The technological progress over the past decades has made it possible to generate, to collect, and to store new intraday high-frequency time series datasets that are widely available along with the “old” low-frequency data. Indeed, the economic activity occurs in real time and the economic and financial transactions are frequently recorded instantaneously, while the traditional time series data are available at a quarterly, monthly, or sometimes daily frequencies. Ignoring the high-frequency nature of the data leads to the loss of the information through the temporal aggregation and makes it impossible to quantify the economic activity in real time. At the same time, combining the low and the high-frequency datasets allows obtaining more refined measures of the economic activity that can be used subsequently to inform market participants and to guide policies.

In this article, we introduce a novel high-dimensional mixed-frequency instrumental variable (IV) regression suitable for the datasets recorded at different frequencies. The model connects a low-frequency dependent variable to endogenous covariates sampled from a continuous-time stochastic process. Alternatively, the regressor might be sampled from a continuous-space stochastic process encountered in the spatial data analysis or any other stochastic process indexed by the continuum. This leads to the high-dimensional IV regression with a large number of endogenous regressors.

The high-dimensional mixed-frequency IV regression features several remarkable properties. First, we show that it is possible to identify and to estimate accurately the high-dimensional slope parameter leveraging on a low frequency IV. In contrast, the point identification in the (high dimensional) linear IV regression typically relies on the order condition postulating that the number of IVs should be at least as large as the number of endogenous regressors. Second, the mixed-frequency IV regression can handle arbitrary large number of endogenous covariates relatively to the sample size without relying on approximate sparsity condition and restrictive tail conditions. Such a remarkable property is possible due to the continuous-time structure of the regressor and the slope parameter. Continuous-time structures is one of the “blessings of dimensionality” along with the concentration of measure and the extreme-value theory according to Donoho (2000). These properties distinguish our model from the ridge IV regression; see Carrasco (2012) or the high-dimensional IV regression of Belloni, Chen, Chernozhukov, and Hansen (2012).

The high-dimensional mixed-frequency IV regression is an example of ill-posed inverse problem in the sense that the map from the distribution of the data to the slope parameter is not continuous. As a result, we need to introduce some amount of regularization to smooth out the discontinuities and to obtain a consistent estimator. The concept of regularization originates from the mathematical literature on ill-posed inverse problems; see Tikhonov (1943, 1963) and an excellent review article of Carrasco, Florens, and Renault (2007) for introduction further references in econometrics. In this article, we focus on the Tikhonov regularization and establish its statistical properties with weakly dependent data. In contrast to well-posed nonparametric problems, the estimation accuracy of the continuous-time slope parameter depends both on its regularity as well as on the regularity of a certain integral operator.
Our empirical application extends the classical IV estimation of the supply and the demand equations, cf. Wright (1928), to the real-time spot markets. We collect a new dataset using publicly available data and estimate the real-time price elasticity of supply for the Australian electricity spot market. To that end, we leverage on the daily temperature as an IV that shifts the demand curve and is exogenous for supply shocks. The temperature is a valid IV since the electricity demand increases in hot and cold times due to cooling and heating needs and it is plausible that the temperature does not affect the supply through other than the demand channels. Our empirical results reveal that the supply of electricity is relatively inelastic and that it may be heterogeneous throughout the day.

1.1. Contribution and Related Literature

This article connects several strands of the literature. First, following Ghysels, Santa-Clara, and Valkanov (2004), Ghysels, Sinko, and Valkanov (2007), and Andreou, Ghysels, and Striaukas (2021a,b), there is an increasing interest in datasets sampled at different frequencies in the empirical practice. Most of this literature, with a notable exception for Khalaf et al. (2021), is largely focused on forecasting problems with mixed-frequency data and does not consider the structural econometric modeling with the IV approach. The mixed-frequency data typically lead to high-dimensional problems and the dimensionality is controlled using tightly parameterized weight functions; see also Foroni, Marcellino, and Schumacher (2015) for the unrestricted mixed-frequency data models. The model considered in the present article has the following features: (i) we introduce a novel IV regression suitable for the data sampled at mixed frequencies and the structural econometric modeling; (ii) we do not rely on a particular parameterization of the weight function; (iii) our high-frequency data are generated from the endogenous continuous-time stochastic process, which is taken into account in our theory.

Second, we build on insights from literature on the Tikhonov regularization of ill-posed inverse problems in econometrics, see Carrasco, Florens, and Renault (2007), Gagliardini and Scaillet (2012), and Carrasco, Florens, and Renault (2014) for comprehensive surveys, and the functional linear IV regression, see Florens and Van Bellegem (2015), Benatia, Carrasco, and Florens (2017), and Babii (2020). In contrast to this literature, we show that it is possible to achieve identification and to estimate accurately the slope parameter relying on a single instrumental measured at a low-frequency only. The structure of our model is also qualitatively different and leads to the conditional expectation operator that was not previously encountered in the ill-posed inverse problems literature.

Finally, following the influential work of Belloni, Chernozhukov, and Hansen (2013) and Belloni, Chen, Chernozhukov, and Hansen (2012), there is an increasing interest in the estimation and inference with high-dimensional datasets in econometrics; see Belloni et al. (2018) for an excellent introduction and further references. In particular, Belloni, Chen, Chernozhukov, and Hansen (2012) proposed to use the LASSO to address the problem of many instruments and the non-parametric series estimation of the optimal instrument. Our mixed-frequency IV regression is qualitatively different from the above models and does not impose the approximate sparsity on the high-dimensional slope coefficients; see Babi, Ghysels, and Striaukas (2021a,b), for approximately sparse mixed-frequency time series regressions. The problem of the optimal instrument is more challenging in our nonparametric setting and is left for future research; see Florens and Sokullu (2018) for some steps in this direction. Last, we also consider the mixed-frequency IV regression with covariates via a time-series version of double/debiased machine learning approach of Chernozhukov et al. (2018). The article is organized as follows. In Section 2, we present the mixed-frequency IV regression, illustrate several economic examples, and discuss the identification issues. In Section 3, we consider the Tikhonov-regularized estimator and develop the large sample theory for the weakly dependent time series data. In particular, we discuss the large sample results for the model with and without covariates (via double/debiased machine learning) and study the infill asymptotics and inference. We report on a Monte Carlo study in Section 4 which provides further insights about the validity of asymptotic analysis in finite samples typically encountered in empirical applications. Section 5 presents an empirical application to the estimation of real-time supply elasticities and Section 6 presents conclusions. All proofs appear in the appendix and the supplementary material.

1.2. Notation

We use $L^2(S)$ to denote the space of functions on $S \subset \mathbb{R}^d$, square-integrable with respect to the Lebesgue measure. We endow the space $L^2(S)$ with the natural inner product $(\beta, \gamma) = \int_S \beta(s) \gamma(s) ds$ and the norm $\|\beta\| = \sqrt{(\beta, \beta)}$ for all $\beta, \gamma \in L^2(S)$. For a random variable $W$, we denote $L^2(W) = \{ f : \mathbb{E}|f(W)|^2 < \infty \}$. Any vector $a \in \mathbb{R}^m$ should be considered as a column-vector and can be written as $a = (a_j)_{1 \leq j \leq m}$. For a bounded linear operator $K : \mathcal{E} \rightarrow \mathcal{F}$ between the two Hilbert spaces $\mathcal{E}$ and $\mathcal{F}$, let $\|K\| = \sup_{\|\theta\|_1 \leq 1} \|K\theta\|$ denote its operator norm. Let $\sigma(K^*K)$ denote the spectrum of the corresponding self-adjoint operator $K^*K$. The $m \times T$ matrix $A$ is written by enumerating all its elements $A = (A_{ij})_{1 \leq j \leq m}$. If $m = T$, then we simply write $A = (A_{ij})_{1 \leq i \leq m}$. We use

$$C^0_{[0,1]} = \left\{ f : [0,1] \rightarrow \mathbb{R} : \max_{k \leq |x|} \|f^{(k)}\|_\infty \leq L, \sup_{x \neq x'} \frac{|f^{(k)}(x) - f^{(k)}(x')|}{|x - x'|^{k+1}} \leq L \right\}$$

to denote the space of H"older continuous functions with parameters $\kappa, L > 0$. For two sequences $(a_n)_{n \in \mathbb{N}}$ and $(b_n)_{n \in \mathbb{N}}$, we write $a_n \lesssim b_n$ if $a_n = O(b_n)$. Similarly, we write $X_n \lesssim P Y_n$ if $X_n = O_P(Y_n)$. Lastly, for $a, b \in \mathbb{R}$, we put $a \wedge b = \min\{a, b\}$.

2. Mixed-Frequency IV Regression

The purpose of this section is to introduce the mixed-frequency IV regression and to discuss the identification strategy.

2.1. The model

Econometrician observes \{(Y_t, Z_t(s), W_t) : t = 1, \ldots, T, j = 1, \ldots, m\}, where \(Y_t \in \mathbb{R}\) is a low-frequency dependent variable, \(Z_t(s)\) is a realization of a real-valued continuous-time stochastic process \(Z_t = \{Z_t(s) : s \in S \subset \mathbb{R}^d\}\), and \(W_t \in \mathbb{R}^d\) is a vector of low-frequency IVs. The dimension \(d\) depends on the case specific, for example, if \(d = 1\), then \(S \subset \mathbb{R}\) can be interpreted as a time index (time series). More generally, if \(d = 2\), then \(S \subset \mathbb{R}^2\) can be a geographical location (spatial process), and if \(d = 3\), then \(S \subset \mathbb{R}^3\) denotes the space and the time dimension (spatio-temporal process). Regardless of the dimension \(d\), we always refer to \(Z_t\) as a continuous-time stochastic process. The number of high-frequency observations \(m\) is left unrestricted and can (potentially) be much larger than the sample size \(T\). The continuous-time mixed-frequency IV regression is

\[
Y_t = \int_S \beta(s)Z_t(s)ds + U_t, \quad \mathbb{E}[U_t|W_t] = 0, \quad t = 1, \ldots, T.
\]

Alternatively, if we start from the linear model \(Y = \Phi(Z) + U\), where \(\Phi : L^2(S) \to \mathbb{R}\) is a continuous linear functional, then by the Riesz representation theorem, we can always write \(\Phi(Z) = \langle \beta, Z \rangle\) for a unique slope parameter \(\beta \in L^2(S)\).

In practice, since the process \(\{Z_t(s) : s \in S\}\) is sampled at discrete time points \(\{Z_t(s) : j = 0, 1, \ldots, m\}\), there will be additional discretization bias vanishing under the infill asymptotics. It is worth stressing that the discretization of the continuous-time model leads to a consistent definition of regression slopes across different frequencies. See Sims (1971) and Geweke (1978).

The mixed-frequency IV regression is appropriate whenever we have endogeneity in the sense that there exists some set \(A \subset S\) of positive Lebesgue measure such that \(\mathbb{E}[U_t|Z_t(s)] \neq 0\forall s \in A\). The following two examples describe the appropriate empirical settings.

**Example 2.1 (Real-time price elasticities).** Spot markets operate in real time with commodities traded for the immediate delivery. Let \(Y^S_t\) and \(Y^D_t\) be the quantities supplied and demanded on a day \(t\). Consider a structural model

\[
Y^S_t = \langle \beta, Z_t \rangle + U_t, \quad Y^D_t = \langle \gamma, Z_t \rangle + V_t,
\]

where \(s \mapsto Z_t(s)\) is the price path on a day \(t\), \(\beta\) is the real-time price elasticity of supply, and \(\gamma\) is the real-time price elasticity of demand. In the market equilibrium \(Y^S_t = Y^D_t\), implying \(U_t = V_t = \langle \gamma - \beta, Z_t \rangle\). Therefore, the supply shock \(U_t\) will be correlated with the price \(Z_t\). The mixed-frequency IV regression can be used to identify and to estimate the real-time elasticities of supply/demand, extending the classical linear IV regression to the continuous-time markets, cf. Wright (1928).

**Example 2.2 (Intraday liquidity).** In the equilibrium of a seminal Kyle (1985) model, \(Y_t\) is a daily price change of an asset \(t\), \(Z_t(s)\) is an order flow imbalance on a day \(t\) at time \(s\), and \(1/\beta\) is a liquidity parameter. The liquidity parameter quantifies the sensitivity of the market price to the imbalance between the supply and the demand. Endogeneity comes from the strategic behavior of informed traders who are likely to distribute orders over time to minimize the impact on prices and the market equilibrium.

2.2. Identification

To simplify the notation, in this section, we suppress the dependence of \((Y_t, Z_t, W_t)\) on \(t\) and write \((Y, Z, W)\), which is well-justified under stationarity. The mixed-frequency IV regression becomes

\[
Y = \int_S \beta(s)Z(s)ds + U, \quad \mathbb{E}[U|W] = 0.
\]

The identification in the linear IV regression relies on the uncorrelatedness between the IV and the unobservables, that is, \(\mathbb{E}[U|W] = 0\), and the rank condition. The rank condition requires in turn that the number of the IV matches the dimension of the endogenous covariate. In our settings, the endogenous covariate is a high-dimensional realization of a continuous-time stochastic process, hence, one would need a high-dimensional IV to identify \(\beta\).

In contrast, our identification strategy relies on the mean independence exogeneity condition, \(\mathbb{E}[U|W] = 0\). Assuming that the order of the integration can be interchanged, this condition implies

\[
h(w) \triangleq \mathbb{E}[Y|W = w] = \int_S \beta(s)\mathbb{E}[Z(s)|W = w]ds \triangleq (L\beta)(w),
\]

where \(L : L^2(S) \to L^2(W)\) is an integral operator mapping the unknown slope parameter \(\beta\) to the conditional mean function \(h\). Equation (1) is an example of the Fredholm integral equation of Type I, solving which is typically an ill-posed problem in the sense that the inverse map from \(h\) to \(\beta\) is discontinuous; see Carrasco, Florens, and Renault (2007).

Our identification strategy relies on the linear completeness property of the distribution of \((Z, W)\). We say that the stochastic process \(Z \in L^2(S)\) is linearly complete for \(W \in \mathbb{R}^d\) if for all \(b \in L^2(S)\) with \(\mathbb{E}[(Z, b)|W] < \infty\), we have

\[
\mathbb{E}[(Z, b)|W] = 0 \implies b = 0.
\]

We assumed this condition throughout the article:

**Assumption 2.1.** The stochastic process \(Z\) is linearly complete for \(W\).

It is worth mentioning that the linear completeness is significantly weaker than the nonlinear completeness condition typically used in the nonparametric IV literature. Indeed, the latter would require that \(\mathbb{E}[\phi(Z)|W] = 0 \implies \phi = 0\) for nonlinear functions of \(Z\) as well. Moreover, it may be relaxed if additional structure is imposed on \(\beta\); see Babii and Florens (2020b). The linear completeness is a generalization of the rank condition imposed in the finite-dimensional linear IV regression and requires that the operator \(L\) is injective. Consider another injective operator \(M : L^2_W \to L^2(V), V \subset \mathbb{R}^p\) such that \((M\phi)(u) = \mathbb{E}[\phi(W)\Psi(u, W)]\) for some square-integrable
function of the IV $\Psi$. Applying $M$ to both sides of Equation (1) leads to
\[
r(u) = E[Y\Psi(u, W)] = \int_S \beta(s)E[Z(s)\Psi(u, W)]ds = (K\beta)(u),
\]
where $r = MH$ and $K = ML : L^2_W \to L^2(V)$ is a new operator. It is more convenient to estimate the slope parameter $\beta$ using the continuum of moment restrictions in Equation (2), since it does not involve conditional expectations, nonparametric estimation of which introduces additional tuning parameters; see Carrasco and Florens (2000) and Carrasco et al. (2007) for the estimation of the finite-dimensional parameters characterized by a continuum of moment conditions. At the same time, Equation (2) has the same identifying power as Equation (1) provided that the operator $M$ is injective. A large class of compactly supported instrument functions that ensure the injectivity of $M$ is characterized in Stinchcombe and White (1998); see also an earlier work of Bierens (1982) who developed consistent specification tests and the work of Dominguez and Lobato (2004) and Lavergne and Patilea (2013) who developed estimators of finite-dimensional parameters based on the Bierens-type trick. Our default recommendation is the logistic CDF, $\Psi(u, W) = 1/(1 + \exp(-u^T W))$ with $u \in V \subset \mathbb{R}^{d+q}$, $W = (1, W^T)^T$, where $V$ is a compact set with nonempty interior.

3. Tikhonov Regularization

In this section, we introduce the Tikhonov-regularized estimator of the slope parameter $\beta$ and study its large sample properties for time series data.

3.1. Estimator

Our objective is to estimate the slope parameter $\beta$ using the continuum of moment conditions in Equation (2), which requires inverting the operator $K$. Note that the integral operator $K$ has the kernel function $k(s, u) \triangleq E[Z(s)\Psi(u, W)]$, which is typically square-integrable. Consequently, the operator $K$ is compact and its generalized inverse is not continuous; see Carrasco, Florens, and Renault (2007). The operator inversion problem is amplified by the fact that $r$ and $K$ are unobserved and have to be estimated from the data. In this article, we focus on the Tikhonov-regularized estimator of $\beta$.

To describe the estimator, let $(Y_t, Z_t, W_t)_{t=1}^T$ be a stationary sample. The operator $K$ and the function $r$ are estimated with sample averages
\[
\hat{r}(u) = \frac{1}{T} \sum_{t=1}^{T} Y_t \Psi(u, W_t),
\]
\[
(K\beta)(u) = \int_k \beta(s)k(s, u)ds,
\]
\[
\hat{k}(s, u) = \frac{1}{T} \sum_{t=1}^{T} Z_t(s)\Psi(u, W_t).
\]

Then the Tikhonov-regularized estimator solves the following penalized least-squares problem:
\[
\hat{\beta} = \arg\min_{b \in L^2(S)} \|Kb - \tilde{r}\|^2 + \alpha\|b\|^2,
\]
where $\alpha > 0$ is a tuning parameter controlling the amount of the regularization and $\|\cdot\|$ denotes the norm on the relevant $L^2$ space. The estimator has a well-known closed-form expression, which resembles the expression of the finite-dimensional ridge regression estimator
\[
\hat{\beta} = (\alpha I + \hat{K}^* \hat{K})^{-1} \hat{K}^* \tilde{r},
\]
where $\hat{K}^*$ is the adjoint operator to $\hat{K}$. It is well-known that the compact self-adjoint operator has a countable, decreasing to zero sequence of eigenvalues. Tikhonov regularization stabilizes the spectrum of the generalized inverse of the operator $\hat{K}^* \hat{K}$, replacing its eigenvalues $1/\hat{\lambda}_j$ by $1/(\alpha + \hat{\lambda}_j)$; see Carrasco, Florens, and Renault (2007) for more details. To compute the adjoint operator, note that for every $\psi \in L^2$, by Fubini’s theorem
\[
(K^* \psi)(s) = \int \psi(s)k(s, u)du.
\]

It is worth mentioning that one could also consider the Tikhonov regularization with Sobolev norm penalty and/or more general spectral regularization schemes, see Carrasco, Florens, and Renault (2007, 2014), Babii and Florens (2020a,b), and references therein.

3.2. Large Sample Theory

To investigate the large sample properties of $\hat{\beta}$, we impose several weak dependence conditions on the underlying stochastic processes. The following definition generalizes the notion of the covariance stationarity to function-valued stochastic processes, see Bosq (2012) for a comprehensive introduction to the statistical theory of stochastic processes in Hilbert and Banach spaces.

Definition 3.1. The $L^2(S)$-valued stochastic process $(X_t)_{t \in \mathbb{Z}}$ is covariance stationary if
\begin{enumerate}[i.]
\item the second moment exists: $\sup_{t \in \mathbb{Z}} E\|X_t\|^2 < \infty$;
\item the mean function is constant over time: $E[X_t(s)] = \mu(s), \forall s \in S$ and $\forall t \in \mathbb{Z}$;
\item the autocovariance function depends only on the distance between observations: $\forall s, u \in S$ and $\forall h, k \in \mathbb{Z}$
\end{enumerate}
\[
\gamma_{h,k}(s, u) = E[(X_{u+h}(s) - \mu(s))(X_{u+k}(u) - \mu(u))],
\]
\[\triangleq \gamma_{h-k}(s, u).
\]

We also need the following notion of the absolute summability of the autocovariance function for $L^2(S)$-valued stochastic processes:

Definition 3.2. The $L^2(S)$-valued covariance stationary process $(X_t)_{t \in \mathbb{Z}}$ has the absolutely summable autocovariance function $\gamma_h$ if
\[
\sum_{h \in \mathbb{Z}} \|\gamma_h\|_1 < \infty,
\]
where $\|\gamma_h\|_1 = \int_S |\gamma_h(s, s)|ds$ denotes the $L_1$ norm on the diagonal of $S \times S$. 

JOURNAL OF BUSINESS & ECONOMIC STATISTICS
The following assumption restricts the dependence structure of the process:

**Assumption 3.1.** \( (u \mapsto U_t\Psi(u, W_t) : t \in \mathbb{Z}) \) and \((s, u) \mapsto Z_s(s)\Psi(u, W_t) : t \in \mathbb{Z} \) are covariance stationary \( L^2 \)-valued stochastic processes with absolutely summable autocovariance functions.

The covariance stationarity in **Assumption 3.1** is a relatively mild condition and is satisfied, in particular, when \((Y_t, Z_t, W_t)_{t \in \mathbb{Z}} \) is strictly stationary. The absolute summability of autocovariances is also a relatively mild condition, typically assumed in the time series analysis. It is worth stressing that the stationarity is imposed on entire trajectories of the processes over \( t \in \mathbb{Z} \), so that on a fixed day \( t \in \mathbb{Z} \), the intraday observations \( Z_t(s) \) for \( s \in \mathbb{S} \) can be nonstationary.

Since the mixed-frequency IV regression model is ill-posed, we also need to quantify the degree of ill-posedness of the operator \( K \) and the regularity of the slope parameter \( \beta \). The following conditions serve this purpose:

**Assumption 3.2.** The slope parameter \( \beta \) belongs to the class

\[ \mathcal{F}(\gamma, R) = \{ b \in L^2(S) : b = (K^* K)^\gamma \psi, \| \psi \| \leq R \} \]

for some \( R > 0 \) and \( \gamma \in (0, 1] \).

To appreciate this condition, note that if \( \beta = (K^* K)^\gamma \psi \), then \( \psi = (K^* K)^{-\gamma} \beta \). Let \( (\sigma_j, \beta_j)_{j=1}^{\infty} \) be the singular values decomposition of the compact linear operator \( K \); see Carrasco, Florens, and Renault (2007). Then \( \beta = \sum_{j=1}^{\infty} \langle \beta_j, \beta \rangle \beta_j \) and by the Parseval’s identity

\[ \| \psi \|^2 = \sum_{j=1}^{\infty} \frac{|\langle \beta_j, \beta \rangle|^2}{\sigma_j^j}. \]

Therefore, \( \beta = (K^* K)^\gamma \psi \) and \( \| \psi \|^2 \leq R \) in **Assumption 3.2** restrict the regularity of the slope parameter \( \beta \) as measured by how fast the Fourier coefficients \( \langle \beta_j, \beta \rangle_{j=1}^{\infty} \) decrease to zero relatively to the smoothing properties of the operator \( K \) as measured by how fast the singular values \( (\sigma_j)_{j=1}^{\infty} \) decrease to zero and the regularity parameter \( \gamma \not\geq 0 \).

The following result describes the linearization of the Tikhonov-regularized estimator, see the appendix for the proof.

**Theorem 3.1.** Suppose that Assumptions 2.1, 3.1, and 3.2 are satisfied. Then

\[ \hat{\beta}(s) - \beta(s) = \frac{1}{T} \sum_{t=1}^{T} \frac{U_t}{\alpha T} + \frac{\alpha^\gamma}{\alpha^T} + \alpha^\gamma. \]

By Lemma A.1.1 in the appendix, the \( L^2 \) norm of the leading term has size \( O_p(1/\sqrt{T}) \). Consequently, if the regularization parameter \( \alpha \to 0 \) as \( T \to \infty \), we obtain

\[ \| \hat{\beta} - \beta \| \leq \frac{1}{\sqrt{T}} + \frac{\alpha^\gamma}{\alpha T} + \alpha^\gamma. \]

To balance the first and the last (leading) terms, we shall set \( \alpha \sim T^{-1/(2\gamma+1)} \), in which case the \( L^2 \) convergence rate is of order \( O_p(T^{-\gamma/(2\gamma+1)}) \).

### 3.3. Double/Debiased Machine Learning

If additional exogenous covariates \( X_t \in \mathbb{R}^r \) are available, then the object of interest is the covariate-adjusted slope parameter \( \beta \). The mixed-frequency IV regression with covariates is

\[ Y_t = \int_S \beta(s) Z_s(s) ds + m(X_t) + U_t, \quad \mathbb{E}[U_t | W_t, X_t] = 0, \]

where \( m : \mathbb{R}^p \to \mathbb{R} \) is an unknown function. Subtracting \( \hat{Y}_t = Y_t - \mathbb{E}[Y_t | X_t] \) and \( \hat{Z}_t(s) = Z_t(s) - \mathbb{E}[Z_t(s) | X_t] \). This leads to the debiased moment condition, inspired by Robinson (1988) and Chernozhukov et al. (2018),

\[ r^d(u) \equiv \mathbb{E}[\hat{Y}_t \hat{\Psi}(u, W_t, X_t)] = \int_S \beta(s) \mathbb{E}[\hat{Z}_t(s) \hat{\Psi}(u, W_t, X_t)] ds \]

where \( \hat{\Psi}(u, W_t, X_t) = \hat{\Psi}(u, W_t) - \mathbb{E}[\hat{\Psi}(u, W_t) | X_t] \). The moment condition involves three nuisance parameters \( \tau_1(X_t) = \mathbb{E}[Y_t | X_t], \tau_2(s, X_t) = \mathbb{E}[Z_t(s) | X_t], \) and \( \tau_3(u, X_t) = \mathbb{E}[\hat{\Psi}(u, W_t) | X_t] \), and as we shall see under mild conditions the estimation of these parameters does not have the first-order asymptotic effect on the estimated slope parameter \( \hat{\beta} \). This is another manifestation of the celebrated Neyman orthogonality property; see Chernozhukov et al. (2018) and references therein for more details.

Let \( \hat{\tau}_1, \hat{\tau}_2, \hat{\tau}_3 \) be the estimators of nuisance parameters obtained from an auxiliary sample, independent of \( (Y_t, Z_t, W_t, X_t)_{t=1}^T \). If the underlying processes are \( M \)-dependent, the auxiliary sample can be easily obtained via the "sample splitting with a gap." The procedure amounts to splitting the observed time series in three blocks such that the first two blocks are of approximately equal size and the middle block of size \( M \) is removed in order to make the two extreme blocks independent. The \( M \)-dependence could perhaps be relaxed to the middle block having size \( M \to \infty \) as \( T \to \infty \). One could also consider the blockwise \( K \)-fold cross-fitting with blocks of size \( M \) removed between the \( K \) folds. This approach, however, reduces the effective sample size by \( (K-1)M \), which may lead to the significant efficiency loss when \( K > 2 \) and the series is short. Alternatively, in the setting of our empirical application, we could also use an auxiliary sample from a similar independent market. Let

\[ \hat{\beta}^d = (\alpha I + \hat{K}^d)^{-1} \hat{K}^d \hat{r}^d \]

be the debiased Tikhonov-regularized estimator of \( \beta \), where

\[ \hat{r}^d(u) = \frac{1}{T} \sum_{t=1}^{T} (Y_t - \hat{\tau}_1(X_t)) (\hat{\Psi}(u, W_t) - \hat{\tau}_3(u, X_t)), \]

\[ (\hat{K}^d \beta)(u) = \int_S \beta(s) \hat{K}^d(s, u) ds, \]

\[ \hat{K}^d(s, u) = \frac{1}{T} \sum_{t=1}^{T} (Z_t(s) - \hat{\tau}_2(s, X_t)) (\hat{\Psi}(u, W_t) - \hat{\tau}_3(u, X_t)). \]
are the debiased estimators of \( r^d, K^d \), and \( k^d(s, u) = E[Z(s)\Psi(u, W_t, X_t)]. \) Let \( (L^2(P_X), \| \cdot \|_{P_X}) \) and \( (L^2(\lambda \otimes P_X), \| \cdot \|_{\lambda \otimes P_X}) \) be the spaces of square-integrable random elements with respect to the marginal distribution of \( X_t \), denoted \( P_X \), and the product of Lebesgue measure \( \lambda \) on \( V \subset \mathbb{R}^{1+q} \) and \( P_X \). Let \( \mathcal{F}_t = \sigma((U_{r-1}, W_{r-1}, Y_{r-1}, Z_{r-1}, X_r); r \le t) \) denote the filtration for every \( t \in \mathbb{Z} \). The following assumption provides a sufficient set of additional conditions for the model with exogenous covariates.

**Assumption 3.3.** (i) \((Y_t, Z_t, W_t, X_t)_{t \in \mathbb{Z}}\) is covariance stationary and \( \sup_{x} E[|U_1|X_t = x] \le C < \infty \); (ii) \((U_t, \Psi(., W_t, X_t), Y_t, Z_t)_{t \in \mathbb{Z}}\) are martingale difference sequences with respect to \((\mathcal{F}_t)_{t \in \mathbb{Z}}\); and (iii) \( \| \hat{\tau}_t - \tau_t \|_{P_X} = O_p(T^{-1/4}) \) and \( \| \hat{\tau}_k - \tau_k \|_{\lambda \otimes P_X} = O_p(T^{-1/4}) \) for \( k = 2,3 \).

Condition (ii) is assumed for the sake of brevity of proofs and might be relaxed to serially correlated processes under additional assumptions on the long-run variances. Condition (iii) requires the nuisance parameters are estimated at a sufficiently fast rate and is standard in the semiparametric estimation literature. It can be verified for various nonparametric/machine learning estimators. Condition (iii) can also be relaxed to a weaker condition that the product of estimation errors converges at the \( O_p(T^{-1/2}) \) rate, in which case, one of estimators can be consistent at a slower rate provided that the other can offset it; see also Farrell (2015) and Chernozhukov et al. (2018). The following result holds for the debiased Tikhonov-regularized estimator in the model with covariates, see supplementary material for the proof.

**Theorem 3.2.** Suppose that assumptions of Theorem 3.1 are satisfied for \((\hat{Y}_t, \hat{Z}_t, \hat{W}_t)_{t \in \mathbb{Z}}\). Suppose also that Assumption 3.3 holds and \( \sup_{w,x} \| \hat{\Psi}(., w, x) \| ^2 < \infty \). Then \( \hat{\beta}(s) - \beta(s) = \frac{1}{T} \sum_{t=1}^{T} U_t([aI + k^d s^{-1} K^d \Psi(., W_t, X_t)](s) + R_T(s)) \) with \( \| R_T \| \le \frac{1}{\alpha T} + \frac{\alpha T^{1/2}}{\sqrt{\alpha T}} + \alpha^\gamma \).

Note that this asymptotic expansion is similar to the one in Theorem 3.1 with exactly the same order of the remainder term \( R_T \). Note that the proof of Theorem 3.2 relies on the independence of two samples used to compute \( \hat{\beta} \) and \( \hat{\tau}_k \), \( k = 1,2,3 \) and it is beyond the scope of the present article to understand whether this condition can be relaxed.

### 3.4. Infill Asymptotics and Inference

So far we have assumed that the trajectory of the stochastic process \( \{Z_t(s) : t = 1, \ldots, T, s \in \mathcal{S} \} \) is completely observed. In this section, we relax this requirement and derive an approximation to the \( L^2 \) norm of the estimator that will be used subsequently for inference.

#### 3.4.1. Infill Asymptotics

Suppose for simplicity that \( \mathcal{S} = [0,1] \) and that we observe the process at discrete time points, \( \{Z_j(s_j) : t = 1, \ldots, T, j = 1, \ldots, m\} \), where \( 0 = s_0 \le s_1 < s_2 < \cdots < s_m = 1 \). Then the operator

\[
(\hat{K} \phi)(u) = \int_0^1 \phi(s)\hat{k}(u, s)ds,
\]

\[
\hat{k}(u, s) = \frac{1}{T} \sum_{t=1}^{T} Z_t(s)\Psi(u, W_t)
\]

is not available. Instead, we observe its discrete-time approximation for every \( \phi \in C[0,1] \)

\[
(\hat{K}_m \phi)(u) = \sum_{j=1}^{m} \phi(s_j)\hat{k}(s_j, u)\delta_j
\]

with \( \delta_j = s_j - s_{j-1} \). Let \( \Delta_m \equiv \max_{1 \le j \le m} \delta_j \) and let \( \hat{\beta}_m \) be the solution to

\[
(\alpha I + \hat{K}^* \hat{K}_m)\hat{\beta}_m = \hat{K}^* \hat{r}.
\]

To characterize the estimation accuracy of the discretized estimator, we impose additionally the following assumption:

**Assumption 3.4.** (i) \( \mathcal{S} \subset C^\gamma_t[0,1] \) a.s. for some \( \kappa \in (0,1] \) and \( \lambda \in (0, \infty) \); (ii) \( \sup_{w} \| \hat{\Psi}(., w) \|^2 \le \hat{\Psi} < \infty \).

**Assumption 3.4 (i) is satisfied, for example, for the Brownian motion on [0,1] for every \( \kappa \in (0,1/2) \); see (Mörters and Peres 2010, corol. 1.20).** (ii) is satisfied, for example, for instrument functions that are uniformly bounded on compact intervals. The following result takes into account the discretization bias, cf. Theorem 3.1; see the appendix for the proof.

**Theorem 3.3.** Suppose that Assumptions 2.1, 3.1, 3.2, and 3.4 are satisfied. Then for every \( s \in \mathcal{S} \)

\[
\hat{\beta}_m(s) - \beta(s) = \frac{1}{T} \sum_{t=1}^{T} U_t([\alpha I + K^* K]^{-1} K^* \Psi(., W_t)](s) + R_T(s),
\]

where \( \| R_T \| \le \frac{1}{\alpha T} + \frac{\alpha T^{1/2}}{\sqrt{\alpha T}} + \alpha^\kappa + \frac{\Delta_m^\kappa}{\alpha^T} \).

Note that the discretized estimator has two sources of the bias. First, the Tikhonov regularization leads to the bias of order \( O(\alpha^\gamma) \). Second, the since the process \( Z_t(s) \) is sampled discretely, this leads to the additional discretization bias of order \( O_p(\Delta_m^\kappa/\alpha^T) \). The regularization bias vanishes whenever \( \alpha \to 0 \) while the discretization bias vanishes provided that \( \Delta_m^\kappa/\alpha^T \to 0 \), which requires that the process is observed more frequently (infill asymptotics). These conditions, in conjunction with \( \alpha T \to \infty \) as \( \alpha \to 0 \) and \( T \to \infty \), ensure the \( L^2 \) consistency of the discretized estimator. In particular, \( m \) can increase faster than exponentially with the sample size.

#### 3.4.2. Inference

It is worth mentioning that the leading process in the expansion of Theorem 3.3 does not converge weakly in the \( L^2([0,1]) \) space; see Babii (2020), Proposition A.2.1. In fact, the inner products of this process with \( h \) can converge weakly at different rates in \( \mathbb{R} \), depending on the direction \( h \in L^2([0,1]) \), which is not the case for estimators in well-posed nonparametric models. One could use suitable Gaussian and bootstrap approximations for
inference on $\beta$, provided that such approximations are available for serially correlated time series data; see Chernozhukov, Chetverikov, and Kato (2014, 2016). Instead, we focus on deriving the limiting distribution of a suitably normalized upper bound of $\|\hat{\beta}_m - \beta\|^2$, which can be used to construct confidence bands for $\beta$; see Imaizumi and Kato (2020) and references therein. To that end, the following assumption describes several additional conditions:

**Assumption 3.5.** (i) $(U_t\Psi(., W_t))_{t \in \mathbb{Z}}$ is a stationary process with strong mixing coefficients satisfying $\sum_{j=1}^{\infty} \delta_j^{2+\delta} \text{E}|U_t\Psi(., W_t)|^{2+\delta} < \infty$ for some $\delta > 0$; (ii) $\alpha^2 T \to \infty$, $\alpha^2 T / \tau_1^2 \to 0$, and $\Delta m T / \alpha^2 \to 0$ as $\alpha \to 0$ and $T \to \infty$.

The following result holds, see supplementary material for the proof.

**Theorem 3.4.** Suppose that assumptions of Theorem 3.3 and Assumption 3.5 are satisfied. Then

$$\alpha T \|\hat{\beta}_m - \beta\|^2 \leq \frac{1}{4} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} U_t\Psi(., W_t) \right\|^2 + o_P(1) \frac{d}{4} \sum_{j=1}^{\infty} \lambda_j \tau_1^2,$$

where $(Z_j)_{j \geq 1}$ are iid $N(0, 1)$ and $(\lambda_j)_{j \geq 1}$ are eigenvalues of $R : L^2(V) \to L^2(V)$, characterized as

$$\langle Rh, h \rangle = \text{var}(U_1(\Psi(., W_1), h)) + 2 \sum_{k=1}^{\infty} \text{cov}(U_1(\Psi(., W_1), h), U_{1+k}(\Psi(., W_{1+k}), h))$$

for every $h \in L^2(V)$.

The critical values of $\frac{1}{4} \sum_{j=1}^{\infty} \lambda_j \tau_1^2$ can be obtained, for example, using the dependent multiplier bootstrap; see Doukhan et al. (2015). To construct a confidence band for $\beta$, consider the discrete path of the Ornstein–Uhlenbeck process

$$e_{t,T}^* = e^{-1/2 \epsilon_t} e_{t-1,T}^* + \sqrt{1 - e^{-2/\epsilon_t}} e_{t}^*,$$

where $(\epsilon_t)_{t \geq 0}$ is an iid $N(0, 1)$ and $(\epsilon_t)_{t \geq 0}$ is a sequence such that $\int T \to \infty$ and $l_{T}/T \to 0$ as $T \to \infty$. Let $q_{1-r_1}$ be the $1 - r_1$ quantile of

$$\frac{1}{T} \sum_{t=1}^{T} e_{t,T}^* \tilde{U}_t(\Psi(., W_t))$$

conditionally on $(Y_t, Z_t, W_t)_{t \geq 1}$, where $(\tilde{U}_t)_{t \geq 1}$ are residuals. Then following Imaizumi and Kato (2020), compute the confidence band for $\beta$ in $L^2[0, 1]$ as follows:

$$C_{t_1, t_2} = \left\{ \frac{\hat{\beta}_m(t) - q_{1-r_1}}{\sqrt{\tau_2} \alpha T}, \frac{\hat{\beta}_m(t) + q_{1-r_1}}{\sqrt{\tau_2} \alpha T} \right\} : t \in [0, 1].$$

This band has coverage probability $1 - r_1$ and does not cover $t_2 \in (0, 1)$ points. In practice, we can take $t_2 = 0.1$ or $0.05$, paying the price for covering a larger fraction of points $1 - t_2$. Note that, for $\beta \in L^2(I)$ with $I \neq [0, 1]$, one would also need to adjust the band by the Lebesgue measure of the set $I$; see Imaizumi and Kato (2020) for more details.

### 4. Monte Carlo Experiments

In this section, we discuss the numerical implementation of our high-dimensional mixed-frequency IV estimator and study its behavior in finite samples with Monte Carlo experiments.

We use the logistic CDF, $P(u, W) = 1/(1 + \exp(-u^T \hat{W}))$, as an instrument function, which satisfies assumptions of our theory since it is uniformly bounded and real-valued, unlike some other choices, cf. Bierens (1982) and Stinchcombe and White (1998). If the process $Z_t$ is observed at uniformly spaced points $j/m$ with $j = 1, \ldots, m$, we can discretize Equation (3) replacing integration with the Riemann sum

$$\alpha \hat{\beta} + (\hat{Z}^T \Psi \hat{Z} \hat{m} / (3m)^2) = \hat{Z}^T \Psi \hat{y} / T^2,$$

where $\hat{\beta} = (\hat{\beta}(j/m))_{1 \leq j \leq m}, \hat{Z} = (Z_{1}(j/m))_{1 \leq j \leq m}, \Psi = (\Psi(j/m, W_t))_{1 \leq j \leq m}$ and $\hat{I}_T$ is a $T \times T$ identity matrix. Then we compute the estimator as

$$\hat{\beta} = (\alpha \hat{I}_T + (\hat{Z}^T \Psi \hat{Z} \hat{m} / (3m)^2)^{-1} \hat{Z}^T \Psi \hat{y} / T^2.$$

It is worth mentioning that the numerical computation of integrals with quadrature rules might also be possible and that we consider the Riemann summation for simplicity of presentation.

There are 5000 replications in each Monte Carlo experiment. We generate samples of $(Y_t, Z_t, W_t)_{t=1}$ size $T \in \{100, 500, 1000\}$ as follows:

$$Y_t = \int_0^1 \beta(s)Z_t(s)ds + U_t,$$

$$Z_t(s) = k(s, W_t) + \sigma B_t(s),$$

$$W_t = 0.5 + 0.7 W_{t-1} + \epsilon_t,$$

$$k(s, w) = \sqrt{s^2 + w^2},$$

$$\epsilon_t \sim \text{iid} N(0, 1),$$

$$U_t = 0.5 \int_0^1 B_t(s)ds + 0.5 V_t,$$

$$V_t \sim \text{iid} N(0, 1),$$

where $B_t(s) : s \in [0, 1], t = 1, \ldots, T$ are independent Brownian motion, generated independently of all other variables and initiated at iid random draws from $U(-1/2, 1/2)$. The parameter $\sigma \in [0.5, 1]$ represents the noise level. We consider two slope parameters $\beta(s) = -10 \exp(s)$ and $\beta(s) = 10s$ with $s \in [0, 1]$. All continuous-time quantities are discretized at 200 equidistant points.

The integrated bias, variance, and MSE are approximated by the Riemann sum on a grid of 100 equidistant points in $[0, 1]$. Table 1 (appendix) presents results of our Monte Carlo experiments for two different population slope parameters. The mixed-frequency IV estimator behaves according to our asymptotic results. We can see the bias/variance tradeoff—as the regularization parameter $\alpha$ tends to zero, the bias decreases while the variance increases. The optimal choice of the regularization parameter should balance the two. The estimator performs better when the sample size increases and the noise level decreases. We can also see that the linear slope parameter is estimated more accurately. Figure 1 (appendix) summarizes graphically the outcome of Monte Carlo experiments for $\alpha = 10^{-6}$. The
shaded gray area represents the pointwise interval between the 2.5% and 97.5% quantiles of the estimator across 5000 replications. Overall, the mixed-frequency IV estimator demonstrates excellent performance across different specifications.

It is worth stressing that since the stochastic $Z$ is observed at $m = 200$ time points, the number of endogenous regressors exceeds the sample size when $T = 100$. In this case, the conventional IV estimator does not exist. At the same time, the naive generalization of the ridge regression and the LASSO are also not appropriate in our setting. The ridge regression would typically require $m/T \to 0$, see Carrasco, Florens, and Renault (2007). The LASSO would require the approximate sparsity, somewhat stronger weak dependence conditions, and $m^{1/\kappa}/T^{1-1/\kappa} \to 0$, where $\kappa$ measures tails and weak dependence; see Babii, Ghysels, and Striaukas (2021b).

### 5. Real-Time Elasticity of Electricity Supply

At the beginning of the 1990s, electricity markets around the world were vertically integrated industries with prices set by regulators. Over the last 30 years, major countries experienced deregulation. Today, electricity is often sold at competitive spot markets where prices are determined according to the laws of supply and demand. Elasticities of supply and demand summarize the behavior of energy producers and consumers, inform market participants, and play an important role in the policy design, forecasting, and energy planning. The real-time elasticity of supply contains a piece of important information on seller’s response to the intraday price fluctuations.

Most of the electricity in Australia is generated, sold, and bought at the National Electricity Market (NEM), which is one of the largest interconnected electricity systems in the world. The NEM started operating as a wholesale spot market in December 1998. It supplies about 200 terawatt-hours of electricity to around 9 million customers each year reaching $16.6$ billion of trades in 2016–2017. The supply and demand come from over 100 competitive generators and retailers participating in the market and are matched instantaneously in real time through a centrally coordinated dispatch process. Generators offer to supply a fixed amount of electricity at a specific time in the future and can resubmit subsequently the offered amount and price if needed. The Australian Energy Market Operator (AMEO) decides which generators will produce electricity to meet the demand in the most cost-efficient way.

We construct a new dataset using publicly available data from the AEMO and the Australian Bureau of Meteorology for the New South Wales in 1999–2018. The central pieces of the dataset are the daily aggregate quantities of the electricity sold at the spot market, intraday high-frequency prices measured each half an hour, and the average daily temperatures. The high-dimensional mixed-frequency IV regression model is

$$
\log Q_t = \int_0^{24} \beta(s) \log P_t(s) ds + U_t, \quad E[U_t|W_t] = 0,
$$

where $Q_t$ is the quantity sold on a day $t$, $P_t(s)$ is the price at time $s$ on a day $t$, and $W_t$ is an IV. To estimate the supply elasticity, we use the average daily temperature as an IV. We expect that the temperature is a valid demand shifter since the electricity demand increases in hot and cold times due to cooling and heating needs. Since the observed prices are measured with half an hour intervals, the regression equation is discretized as

$$
\log Q_t = 0.5 \sum_{j=1}^{48} \beta(s_j) \log P_t(s_j) + U_t,
$$

where $s_j = 0.5j$ with $j = 1, 2, \ldots, 48$.

Figure 2 (appendix), panels (a) and (c), show the histogram of the natural logarithm of equilibrium quantities and the boxplot with equilibrium prices plotted against the hour. Panels (b) and (d) display the histogram of the temperature and the scatterplot with quantities plotted against the temperature. Marginal distributions seem to be well-behaved. The price series seems to be nonstationary during the day with the median price peaking in the evening and plummeting during the night. There is also more volatility in the price in the evening. Note that our assumptions do not rule out intraday nonstationarities.

To compute the estimator, following Feve and Florens (2010), we estimate the regularization parameter minimizing

$$
\text{RSS}(\alpha) = \alpha^{-1} \| \hat{K}_{\alpha} - \hat{r} \|^2,
$$

where $\hat{r}_0 = (\alpha I + \hat{K}^*\hat{K})^{-1}\hat{K}^*\hat{r}$. In our case, the minimum is reached at $\alpha^* = 2.16 \times 10^{-3}$ as be seen from Figure 3, panel (a).

Figure 3 (appendix), panel (b) displays the estimated intraday elasticity of supply using the high-dimensional mixed-frequency IV regression. We find that depending on the hour,
Figure 1. Monte Carlo experiments. Summary of Monte Carlo experiments for two functional forms: $\beta(s) = -10 \exp(s)$ (top figures) and $\beta(s) = 10 s$ (bottom figures) with the regularization parameter $\alpha = 10^{-6}$, different noise level $\sigma$, and sample size $T$. Population parameter $\beta$ (solid line), average estimates $\hat{\beta}$ (dashed line) and 95% pointwise confidence band (gray area). The number of experiments is 5000.

The point estimates of price elasticity of supply range between 0.135 and 0.165. The supply appears to have the real-time price elasticity of a similar order of magnitude as the demand, cf. Patrick and Wolak (2001) who found that the real-time demand
elasticieties between 0 and -0.27 for 5 industrial sectors in UK. The relatively inelastic supply may probably be attributed to the fact that the market participants are allowed to hedge financial risks and the difficulty to adjust the electricity production in real time. The figure also displays the 90% confidence, where we set $\tau_2 = 0.1$ and take $l_T = T/5$. The confidence band reveals that the elasticity of supply is in the range $[0.1, 0.2]$ most of the time. Setting smaller $\tau_2$ allows us to cover more than 90% of points, however, the band becomes wider in this case. We also find that smaller values of $l_T$ also increase the width of the band. Finally, since the confidence band based on Theorem 3.4 is already conservative, we do not perform additional undersmoothing; see Babii (2020) for a discussion of the data-driven undersmoothing rule.

Figure 2. Data – The figures summarize the distribution of the data. Panels (a) and (b) show the histograms of daily quantities (in logs) and temperatures. Panel (c) shows the boxplot with the intraday distribution of prices (in logs) and illustrates the intraday variation in electricity prices. Panel (d) shows the scatterplot with the daily quantity (in logs) and temperature and indicates that the temperature is a plausible IV for the quantity demanded.
6. Conclusions
This article introduces a novel high-dimensional mixed-frequency IV regression and contributes to the growing literature on high-dimensional and mixed-frequency data. We show that the slope parameter of the high-dimensional endogenous regressor can be identified and accurately estimated leveraging on an IV observed at a low-frequency only. We develop the large-sample theory for the Tikhonov-regularized estimator with time series data. The mixed-frequency IV estimator has a closed-form expression and is easy and fast to compute numerically. Our statistical analysis does not restrict the number of high-frequency observations of the process and can handle the number of covariates increasing with the sample size even faster than exponentially. We also cover inference and the model with covariates via a time-series extension of the double/debiased machine learning.

In our empirical application, we estimate the real-time price elasticity of supply at the Australian electricity spot market. We find that the supply is relatively inelastic and that it may be heterogeneous throughout the day. To conclude, we shall note that our identification strategy with a low-frequency IV can also be applied to the IV model of Benatia, Carrasco, and Florens (2017) with a high-frequency-dependent variable.

Supplementary Materials
The Supplementary Material contains detailed proofs of Theorems 3.2 and 3.4.

Acknowledgments
This article is a substantially revised Chapter 2 of my Ph.D. thesis. I’m deeply indebted to my advisor Jean-Pierre Florens as well as Eric Gautier, Ingrid van Keilegom, and Timothy Christensen for helpful discussions and suggestions. I thank to participants at the Conference on Inverse Problems in Econometrics at Northwestern University, the ENTER seminar at Tilburg University, the 48èmes Journées de Statistique de la SFdS, the 3rd ISNPS Conference, and the Recent Advances in Econometrics Conference at TSE Toulouse as well as Aleksandra Babii, David Benatia, Bruno Biais, Pascal Lavergne, and Mario Rothschild. I’m also grateful to anonymous referees whose comments helped me to improve significantly the article. All remaining errors are mine.

Appendix. Proofs
To prove Theorem 3.1, we need an additional lemma that bounds the expected norm of the sample mean of a covariance stationary zero-mean $L^2(S)$-valued stochastic process $(X_t)_{t \in Z}$ by the norm of its autocovariance function $\gamma_h$.

Lemma A.1.1. Suppose that $(X_t)_{t \in Z}$ is a zero-mean covariance stationary process in $L^2(S)$ with absolutely summable autocovariance function $\sum_{h \in Z} \|\gamma_h\|_1 < \infty$, where $\|\gamma_h\|_1 = \int_S \|\gamma_h(s, s)\| ds$. Then

$$E \left\| \frac{1}{T} \sum_{t=1}^{T} X_t \right\|^2 \leq \frac{1}{T} \sum_{h \in Z} \|\gamma_h\|_1.$$

Proof. We have

$$E \left\| \frac{1}{T} \sum_{t=1}^{T} X_t \right\|^2 = \frac{1}{T^2} E \left\{ \sum_{t=1}^{T} X_t, \sum_{k=1}^{T} X_k \right\} = \frac{1}{T^2} \sum_{t=1}^{T} \int_S E[X_t(s)X_k(s)] ds = \frac{1}{T^2} \sum_{t=1}^{T} \int_S \gamma_{T-t} ds \leq \frac{1}{T} \sum_{h \in Z} \int_S \|\gamma_h(s, s)\| ds,$$

where the second line follows by the bilinearity of the inner product and Fubini’s theorem and the third under the covariance stationarity.

The following lemma allows controlling estimation errors appearing in the proof of Theorem 3.1 in terms of more primitive quantities.

Lemma A.1.2. Suppose that $\hat{k}, \hat{k}, \hat{r}, r, \beta$ are square-integrable. Then

$$E \left\| \hat{K} - K \right\|^2 \leq E \left\| \hat{k} - k \right\|^2.$$
The proof is based on the following decomposition:
\[
\hat{\beta} - \beta = (aI + K^*K)^{-1}K^*(\hat{\beta} - \hat{K}\beta) + R_1 + R_2 + R_3 + R_4 \tag{A.1}
\]
with
\[
R_1 = \left[(aI + \hat{K}K)^{-1}\hat{K} - (aI + K^*K)^{-1}K\right](\hat{\beta} - \hat{K}\beta)
\]
\[
R_2 = a(aI + \hat{K}K)^{-1}\hat{K}(K - K)\alpha(1 + K^*K)^{-1}b
\]
\[
R_3 = a(aI + \hat{K}K)^{-1}(K^* - K)\alpha(1 + K^*K)^{-1}b
\]
\[
R_4 = (aI + K^*K)^{-1}K^*\beta - \beta.
\]

To see that this decomposition holds, note that
\[
R_1 = a(aI + \hat{K}K)^{-1}\hat{K} - (aI + K^*K)^{-1}K\beta
\]
\[
=a(aI + \hat{K}K)^{-1}\left[(aI + \hat{K}K) - (aI + K^*K)\right]\beta
\]
\[
= a(aI + \hat{K}K)^{-1}0 - a(aI + \hat{K}K)^{-1}\beta
\]
\[
\beta \leq \left[1 - a(aI + \hat{K}K)^{-1}\right] + a\left[aI + K^*K\right]^{-1}I \beta
\]
\[
= (aI + \hat{K}K)^{-1}\hat{K}\beta - (aI + K^*K)^{-1}K\beta.
\]

Note also that
\[
\hat{\beta} - \hat{K}\beta = \frac{1}{T} \sum_{t=1}^{T} U_t \Psi(., W_t).
\]

Therefore, if we can show the desired order for \(E\|\hat{\beta} - \beta\|^2\), the conclusion of the theorem would follow. Under Assumption 3.1 by Lemma A.1.1
\[
E\|\hat{\beta} - \beta\|^2 = \frac{1}{T} \sum_{t=1}^{T} \left(Y_t \Psi(., W_t) - E[Y_t \Psi(., W_t)]\right)^2
\]
and
\[
E\|\hat{K}\beta - K\beta\|^2 = \frac{1}{T} \sum_{t=1}^{T} \left(Z_t \Psi(., W_t) - E[Z_t \Psi(., W_t)]\right)^2
\]

and
\[
E\|\hat{K}\beta - K\beta\|^2 \leq 2E\|\hat{\beta} - \beta\|^2 + 2\|\beta\|^2 E\|\hat{K} - K\|^2.
\]

Since
\[
\|\hat{R}_T\| \leq \|\hat{R}_1\| + \|\hat{R}_2\| + \|\hat{R}_3\| + \|\hat{R}_4\|.
\]
it is sufficient to control each of the four terms separately.

The fourth term is a regularization bias and its order follows directly from Assumption 3.2 and the isometry of the functional calculus
\[
\|\hat{R}_4\| \leq \left\| (a + \hat{K}K)^{-1}K^*K - I \right\| \beta
\]
\[
\leq \left\| I - (a + \hat{K}K)^{-1}K^*K \right\| \beta
\]
\[
= \sup_{\lambda \in (K^*K)} \left\| 1 - \frac{\beta}{\alpha + \lambda} \right\| \lambda^{1/2}
\]
\[
= \sup_{\lambda \in (K^*K)} \left\| \frac{\lambda^{1/2}}{\alpha + \lambda} \right\| \alpha
\]

We can have two cases depending on the value of \(\gamma > 0\). For \(\gamma \in (0, 1)\), the function \(\lambda \rightarrow \lambda^\gamma/(\alpha + \lambda)\) admits maximum at \(\lambda = a\gamma/(1 - \gamma)\). For \(\gamma \geq 1\), the function \(\lambda \rightarrow \lambda^\gamma/(\alpha + \lambda)\) is strictly increasing on \([0, \infty)\), attaining maximum at the end of the spectrum \(\lambda = \|K^*K\|\). Therefore, since \(\gamma^\gamma(1 - \gamma)^{1/2} \leq 1, \gamma \in (0, 1)\), we have
\[
\sup_{\lambda \in (K^*K)} \left\| \frac{\lambda^{1/2}}{\alpha + \lambda} \right\| \alpha \leq \frac{\|K^*K\|^{1/2}}{\alpha^{1/2}} \leq \frac{\|K^*K\|^{1/2}}{\alpha^{1/2}}.
\]

This gives \(\|\hat{R}_4\| \leq a\gamma^\gamma\).

Next, note that \(\hat{K}\) is a finite-rank operator, and hence, compact. Therefore,
\[
\|\hat{R}_2\| \leq \left\| (aI + \hat{K}K)^{-1}\hat{K}\beta \right\| \leq \sup_{\lambda \in (K^*K)} \left\| \frac{\lambda^{1/2}}{\alpha + \lambda} \right\| \alpha \beta \leq \frac{\|K^*K\|^{1/2}}{\alpha^{1/2}} \beta.
\]

Lastly, similar computations yield
\[
\|\hat{R}_1\| \leq \left\| (aI + \hat{K}K)^{-1}\hat{K}^* - (aI + K^*K)^{-1}K^* \right\| \beta
\]
\[
= \left\| (aI + \hat{K}K)^{-1}\hat{K}^* - (aI + K^*K)^{-1}K^* \right\| \beta
\]
\[
= \left\| (aI + \hat{K}K)^{-1}(K^* - K)\right\| \beta
\]
\[
\leq \frac{\|K^*K\|^{1/2}}{\alpha^{1/2}} \frac{\beta}{\alpha^{1/2}} \leq \frac{\|K^*K\|^{1/2}}{\alpha^{1/2}} \|K^*K\| \beta
\]

Combining all the estimates, we obtain
\[
\|\hat{R}_T\| \leq \frac{1}{\alpha^{1/2}} + \frac{\|K^*K\|^{1/2}}{\alpha^{1/2}} + a\gamma^\gamma.
\]

\[\square\]

Proof of Theorem 3.3. Decompose \(\hat{\beta}_m - \beta = \hat{\beta}_m - \hat{\beta} + \hat{\beta} - \beta\).

By Theorem 3.1, we know that \(\|\hat{\beta} - \beta\| \leq \frac{1}{\alpha^{1/2}} + a\gamma^\gamma\). Consequently, it remains to control the discretization error \(\|\hat{\beta}_m - \hat{\beta}\|\).
To that end, note that if \( \hat{\psi}_m \) solves \((\alpha I + \hat{K}_m \hat{K}^*)\hat{\psi}_m = \hat{r} \), then \( \hat{\beta}_m = \hat{K}^* \hat{\psi}_m \). Therefore, \( \hat{\beta}_m = \hat{K}^* + (\alpha + \hat{K}_m \hat{K}^*)^{-1} \hat{r} \). Next, decompose

\[
\hat{\beta}_m - \hat{\beta} = \hat{K}^* (\alpha I + \hat{K}_m \hat{K}^*)^{-1} \hat{r} - \hat{K}^* (\alpha I + \hat{K}_m \hat{K}^*)^{-1} \hat{r} = \hat{K}^* \left((\alpha I + \hat{K}_m \hat{K}^*)^{-1} - (\alpha I + \hat{K}_m \hat{K}^*)^{-1}\right) \hat{r} = \hat{K}^* (\alpha I + \hat{K}_m \hat{K}^*)^{-1} (\hat{K} - \hat{K}_m) \hat{K}^* (\alpha I + \hat{K}_m \hat{K}^*)^{-1} \hat{r}.
\]

(2.8)

Then

\[
\left\| \hat{\beta}_m - \hat{\beta} \right\| \leq \left\| \hat{K}^* (\alpha I + \hat{K}_m \hat{K}^*)^{-1} \right\| \left\| (\hat{K} - \hat{K}_m) \hat{K}^* \right\| \leq P \frac{1}{\hat{\delta}^2} \left\| (\hat{K} - \hat{K}_m) \hat{K}^* \right\|.
\]

The expression inside of the operator norm is the integral operator such that for every \( \psi \in L^2 \)

\[
(\hat{K} - \hat{K}_m) \hat{K}^* \psi = \int \psi(u) \left( \int \hat{k}(s, v) \hat{k}(s, u) \, ds - \sum_{j=1}^{m} \hat{k}(s_j, v) \hat{k}(s_j, u) \delta_j \right) \, du.
\]

Therefore, by the same computations as in Equation (2.1) and the triangle inequality

\[
\left\| (\hat{K} - \hat{K}_m) \hat{K}^* \right\| \leq \int \hat{k}(s, \cdot) \hat{k}(s, \cdot) \, ds - \sum_{j=1}^{m} \hat{k}(s_j, \cdot) \hat{k}(s_j, \cdot) \delta_j \leq \frac{1}{T^2} \sum_{t=1}^{T} \sum_{k=1}^{T} \left\| \Psi(\cdot, W_t) \right\| \left\| \Psi(\cdot, W_k) \right\| \left| \int Z_t(s) Z_k(s) \, ds - \sum_{j=1}^{m} Z_t(s_j) Z_k(s_j) \delta_j \right| \leq \frac{1}{T^2} \sum_{t=1}^{T} \sum_{k=1}^{T} \left\| \Psi(\cdot, W_t) \right\| \left\| \Psi(\cdot, W_k) \right\| \max_{1 \leq t \leq T} \left| \int Z_t(s) Z_k(s) \, ds - \sum_{j=1}^{m} Z_t(s_j) Z_k(s_j) \delta_j \right| \leq \frac{1}{T^2} \sum_{t=1}^{T} \sum_{k=1}^{T} \left\| \int Z_t(s) Z_k(s) \, ds - \sum_{j=1}^{m} Z_t(s_j) Z_k(s_j) \delta_j \right| \leq \frac{1}{T^2} \sum_{t=1}^{T} \sum_{k=1}^{T} \left\| \Psi(\cdot, W_t) \right\| \left\| \Psi(\cdot, W_k) \right\| \max_{1 \leq t \leq T} \left| \int Z_t(s) Z_k(s) \, ds - \sum_{j=1}^{m} Z_t(s_j) Z_k(s_j) \delta_j \right|.
\]

Under Assumption 3.4 (i)

\[
\left| Z_t(s_j) Z_k(s_j) - Z_t(s_i) Z_k(s_i) \right| \leq \left| Z_t(s_j) - Z_t(s_i) \right| \left| Z_k(s_j) \right| + \left| Z_t(s_i) \right| \left| Z_k(s_j) - Z_k(s_i) \right| \leq 2L^2 |s_j - s_i|,
\]

and whence

\[
\sum_{j=1}^{m} Z_t(s_j) Z_k(s_j) \delta_j - \int Z_t(s) Z_k(s) \, ds = \sum_{j=1}^{m} \int_{s_j-1}^{s_j} \left[ Z_t(s_j) Z_k(s) - Z_t(s_j) Z_k(s_j) \right] \, ds \leq \sum_{j=1}^{m} \int_{s_j-1}^{s_j} \left[ Z_t(s_j) Z_k(s_j) \right] \, ds \leq 2L^2 \sum_{j=1}^{m} \int_{s_j-1}^{s_j} \left| s_j - s_k \right| \, ds \leq 2L^2 \sum_{j=1}^{m} \int_{s_j}^{s_j} \left| s_j - s_k \right| \, ds \leq 2L^2 \sum_{j=1}^{m} \int_{s_j}^{s_j} \left| s_j - s_k \right| \, ds \leq 2L^2 \sum_{j=1}^{m} \int_{s_j}^{s_j} \left| s_j - s_k \right| \, ds \leq 2L^2 \max_{1 \leq j \leq m} \delta_{s_j}.
\]

This shows that \( \left\| \hat{\beta}_m - \hat{\beta} \right\| \leq P \Delta_{n}^m / \alpha^{3/2} \).
Farrell, M. (2015), “Robust Inference on Average Treatment Effects With Possibly more Covariates than Observations,” *Journal of Econometrics*, 189, 1–23. [1475]

Fève, F., and Florens, J. (2010), “The Practice of Non-Parametric Estimation by Solving Inverse Problems: The Example of Transformation Models,” *Econometrics Journal*, 13, S1–S27. [1477]

Florens, J.-P., and Sokulu, S. (2018), “Is There an Optimal Weighting for Linear Inverse Problems,” University of Bristol Working Papers. [1471]

Florens, J.-P., and Van Bellegem, S. (2015), “Instrumental Variable Estimation in Functional Linear Models,” *Journal of Econometrics*, 186, 465–476. [1471,1477]

Foroni, C., Marcellino, M., and Schumacher, C. (2015), “Unrestricted Mixed Data Sampling (MIDAS): MIDAS Regressions With Unrestricted Lag Polynomials,” *Journal of the Royal Statistical Society*, Series A, 178, 57–82. [1471]

Gagliardini, P., and Scaillet, O. (2012), “Tikhonov Regularization for Non-parametric Instrumental Variable Estimators,” *Journal of Econometrics*, 167, 61–75. [1471]

Geweke, J. (1978), “Temporal Aggregation in the Multiple Regression Model,” *Econometrica*, 46, 643–661. [1472]

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004), “The MIDAS Touch: Mixed Data Sampling Regression Models,” UNC Working Paper. [1471]

Ghysels, E., Sinko, A., and Valkanov, R. (2007), “MIDAS Regressions: Further Results and New Directions,” *Econometric Reviews*, 26, 53–90. [1471]

Imaiizumi, M., and Kato, K. (2020), “A Simple Method to Construct Confidence Bands in Functional Linear Regression,” *Statistica Sinica*, 29, 2055–2081. [1476]

Khalaf, L., Kichian, M., Saunders, C., and Voia, M. (2021), “Dynamic Panels With MIDAS Covariates: Nonlinearity, Estimation and Fit,” *Journal of Econometrics*, 220, 589–605. [1471]

Kyle, A. S. (1985), “Continuous Auctions and Insider Trading,” *Econometrica*, 53, 1315–1335. [1472]

Lavergne, P., and Patilea, V. (2013), “Smooth Minimum Distance Estimation and Testing With Conditional Estimating Equations: Uniform in Bandwidth Theory,” *Journal of Econometrics*, 177, 47–59. [1473]

Mörters, P., and Y. Peres (2010): *Brownian Motion*, Cambridge, UK: Cambridge University Press. [1475]

Patrick, R. H., and F. A. Wolak (2001): “Estimating the Customer-Level Demand for Electricity Under Real-Time Market Prices,” NBER Working Paper. [1478]

Robinson, P. M. (1988), “Root-N-Consistent Semiparametric Regression,” *Econometrica*, 56, 931–954. [1474]

Sims, C. A. (1971), “Discrete Approximations to Continuous Time Distributed Lags in Econometrics,” *Econometrica*, 39, 545–563. [1472]

Stinchcombe, M. B., and White, H. (1998), “Consistent Specification Testing With Nuisance Parameters Present Only Under the Alternative,” *Econometric theory*, 14, 295–325. [1473,1476]

Tikhonov, A. N. (1943), “On the Stability of Inverse Problems (in Russian),” *Doklady Akademii Nauk*, 143, 195–198. [1470]

——— (1963), “On the Regularization of Ill-Posed Problems (in Russian),” *Doklady Akademii Nauk*, 153, 49–52. [1470]

Wright, P. G. (1928), *Tariff on Animal and Vegetable Oils*, New York: Macmillan Company. [1471,1472]