LCERPA Working paper No. 2024-7

November 2024

Don't Ruin the Surprise: Temporal Aggregation Bias in Structural Innovations

Stephen Snudden
Wilfrid Laurier
University

Don't Ruin the Surprise: Temporal Aggregation Bias in Structural Innovations*

Stephen Snudden†
Wilfrid Laurier University

May 11, 2023

Updated: November 19, 2024

Abstract

Structural innovations estimated from temporally aggregated data (i.e., sums, averages) are shown to be biased and predictable. Selective sampling is shown to rectify this bias only under specific conditions. A statistical test for temporal aggregation bias is proposed, which reveals that over 70 percent of structural innovations can be predicted using lagged daily information. Applying this test, we find significant mistiming in the shocks to the global crude oil market, and show that economic agents have already responded to the predictable component. This predictability and mistiming challenge the reliability of structural economic conclusions drawn from monthly or quarterly data.

JEL classification: C32, C43, C52, Q47
Keywords: Temporal aggregation, structural econometrics, causal inference, oil prices.

*Thanks to Reinhard Ellwanger and Lutz Kilian for their feedback and suggestions. Thank you also to participants at the 2023 Canadian Economic Association Annual Conference, the USAEE, and the Bank of Mexico for comments.

†Department of Economics, 64 University Ave W, Waterloo, Ontario, N2L 3C7, Canada; ORCID iD 0000-0002-4033-3235; Email: ssnudden@wlu.ca

1 Introduction

Structural economic models are often estimated using temporally aggregated data, such as monthly or quarterly averages of daily observations. However, temporal aggregation transforms the underlying data, making the information set available to econometricians quite different from that of economic agents (Christiano and Eichenbaum, 1987). Herein, we quantify the extent to which this discrepancy leads to structural innovations becoming predictable based on past disaggregated information.

Sufficient conditions are provided, and a method is proposed to quantify the predictability of structural innovations using lagged daily information. It is shown that shocks will always be predictable when temporally aggregated data is used in estimation and the underlying data is persistent. While selective sampling can eliminate this bias, it does so only under specific conditions. The results indicate that estimating econometric models, such as structural vector autoregressions, using monthly or quarterly data should be expected to result in mistimed and predictable structural shocks.

A statistical test for temporal aggregation bias is proposed. Simulations show that using monthly averages of daily data can lead to up to 74% of structural innovations being predictable based on lagged information in autoregressive models. This predictability arises from the mechanical loss of information due to aggregation, and it increases with the persistence of the underlying data.

We apply the test to widely used structural shock estimates for the global crude oil market (Kilian, 2009; Baumeister and Hamilton, 2019). Our findings indicate that up to half of the variation in structural innovations is predictable using daily data from the previous months. Additionally, we identify evidence that some structural innovations follow moving average processes at the monthly frequency, a pattern consistent with temporal aggregation and indicative of model misspecification (Working, 1960; Telser, 1967; Marcellino, 1999). Crucially, we show that economic agents have already responded to the predictable component of these structural innovations.

A key contribution of this paper is to highlight the substantial information loss that occurs when disaggregated daily data is aggregated into monthly or quarterly frequencies. Previous studies, which have largely focused on the aggregation of already aggregated data (e.g., monthly to quarterly), have identified only minor effects (Tiao, 1972; Tiao and Wei, 1976; Wei, 1978; Zellner and Montmarquette, 1971; Abraham, 1982; Rossana and Seater, 1995; Georgoutsos et al., 1998;

Athanasopoulos et al., 2011; Marcellino, 1999). However, high degrees of aggregation, such as total output over days within a quarter or monthly average prices, are commonplace. This daily data, often derived from financial markets, exhibits high persistence, which is known to result in significant information loss for forecasts (see, e.g., Tiao, 1972; Amemiya and Wu, 1972) and model fit (Teles and Sousa, 2017). This paper quantifies the considerable mistiming and predictability of structural shocks introduced when such persistent daily data is temporally aggregated.

This paper contributes to the broader understanding of temporal aggregation (see surveys of Marcellino, 1999; Silvestrini and Veredas, 2008). Our quantitative estimates relate to losses in model fit (e.g., Geweke, 1978; Teles and Sousa, 2017), spurious causality (Wei, 1982; Marcellino, 1999; Breitung and Swanson, 2002), and information loss in forecasting (Amemiya and Wu, 1972; Kohn, 1982; Lütkepohl, 1986). Although parameter mapping from disaggregated to aggregated models has been a central focus in prior research, we show that this mapping becomes a poor approximation when the underlying data is highly persistent. The mapping is more useful for parameterizing our proposed test and determining the extent of shock mistiming.

These findings call into question the validity of structural innovations estimated from temporally aggregated data, a common practice in economic applications. They suggest that economic agents have already observed and responded to the information that economists interpret as new information.

2 Information Loss and Predictability

2.1 Temporal Aggregation

Consider the econometricians' information loss from temporal aggregation. Suppose the data generating process of the daily data on day i in month t is given by is an autoregressive moving average model, ARMA(p, q), with structural innovation $\epsilon_{t,i}$,

$$a(L)y_{t,i} \quad = \quad b(L)\epsilon_{t,i} \quad \text{for } i = 1, 2, ..., n;\ t = 1, 2, ..., T. \tag{1}$$

where L is a lag operator, such that $Ly_{t,i} = y_{t,i-1}$, $b(L) = (1 + \alpha_1 L + \cdots + \alpha_q L^q)$, and $a(L) = (1 - \rho_1 L - \cdots - \rho_p L^p)$. Here, n is the number of daily price observations within a month, and we define $p_{t,0} = p_{t-1,n}$ to transition between months.

This daily data is observed by economic agents and could, for example, represent daily observations in financial markets such as commodity prices or interest rates. Now, suppose that the daily data is ignored, and instead the time series used in estimation is the monthly average data, $\overline{y}_t \equiv \frac{1}{n}\sum_{i=1}^{n} y_{t,i}$. This results in two immediate implications.

First, it is well established that temporal aggregation modifies the structure of the time series (Working, 1960; Telser, 1967). Additional moving average terms are introduced for all ARIMA data generating processes (Brewer, 1973; Weiss, 1984) and for vector autoregressive models (Marcellino, 1999). Methods have been motivated specifically to capture these moving averages (Lütkepohl, 2006; Foroni et al., 2019). Specifically, $\bar{y}_t$ follows an ARMA(p, q^*) process where $q^* = NLI(p + 1 + (q - p - 1)/n)$ and $NLI(\cdot)$ is the next-lowest integer function.[1] Moreover, the coefficient has been modified such that the limiting model for an ARIMA(p, d, q) as $n \to \infty$ is an IMA(d, d) (Tiao, 1972; Stram and Wei, 1986). For stationary data, in the limit as n approaches infinity $\bar{y}_t$ becomes white noise.

The second effect is that temporal aggregation has resulted in a reduction in information for econometricians relative to the information set of economic agents. That is, the disaggregated data always contains more information about the aggregated series than the aggregated series itself (Tiao, 1972; Tiao and Wei, 1976; Kohn, 1982). The loss in information is established, in theory, to result in a reduction of forecast accuracy (Amemiya and Wu, 1972; Lütkepohl, 1986). The loss in information is also known to result in a loss of model fit (for example, Geweke, 1978; Georgoutsos et al., 1998; Breitung and Swanson, 2002; Teles and Sousa, 2017).

Let us now investigate these two effects on the predictability of the shocks from the model estimated at the monthly frequency. For arbitrary n, define the aggregation operator $\omega(L) = 1 + L + L^2 + \cdots + L^{n-1}$. Then, introduce a polynomial, $\beta(L)$, which is such that the product $h(L) = \beta(L)a(L)$ only contains powers of L^n. This polynomial $\beta(L)$ always exists, its coefficients depend on those in $a(L)$, and the degree in L is at most equal to $p(n-1)$ (Brewer, 1973; Weiss, 1984). That is, $h(L)$ is a polynomial of the form $(h_0 L^0 + h_1 L^n + h_2 L^{2n} + \cdots + h_{p(n-1)} L^{p(n-1)})$. We then multiply both sides by $\omega(L)$ and $\beta(L)$ and obtain

$$\begin{aligned} \omega(L)\beta(L)a(L)y_{t,i} &= (1 + \cdots + L^{n-1})(1 + \cdots + \rho^{n-1} L^{p(n-1)})(1 + \cdots + \alpha_q L^q)\epsilon_{t,i} \\ h(L)\bar{y}_t &= (1 + (1 + \rho_1 + \alpha_1)L + \cdots + \alpha_q \rho_p^{n-1} L^{(p+1)(n-1)+q})\epsilon_{t,i}. \end{aligned} \tag{2}$$

where $\bar{y}_t = \omega(L)y_{t,i}$ is the temporal aggregate. Note that $L^n \epsilon_{t,i} = \epsilon_{t-1,n}$ and largest lag operator on

[1] The next-lowest integer function, $NLI(\cdot)$, means that the expression in brackets is rounded down to the next-lowest integer if it is not already an integer.

$\epsilon_{t,i}$ is $(p+1)(n-1)+q$. Equation 2 shows the statistical errors exhibit high-frequency moving average components. The coefficients on the high-frequency moving average terms are non-constant.

The loss of information occurs when only the aggregate data is used in estimation. Approximating the high-frequency moving average terms with moving average terms at the lower frequency incorrectly imposes a constant moving average coefficient on $\epsilon_{t,i}$ within t for each q^*. This adequately fails to capture the non-constant weights on high-frequency moving average components introduced from data transformation. Such a model misspecification implies that innovations from the aggregate model, u_t, exhibit $covar(u_t, L^j y_{t-1,n}) \neq 0 \; \forall \; j \in \{0, 1, ..., p(n-1)+q\}$. Thus, predictability of the structural innovations are expected for all ARMA(p, q) representations of the daily data since it is always the case that $(p+1)(n-1)+q > n-1$, $\forall \; n > 1$ when p, $q > 0$. Further note that when $j > n$, aggregation places non-linear weights on daily observations from more than one lag of the temporally aggregated periods. Hence, temporal aggregation bias implies that the estimates from models relying on temporally aggregated data can be predictable up to several periods in the past.

2.1.1 A Simple Example

For illustrative purposes, consider the case when the data generating process of the daily data is given by is an autoregressive model, AR(p), with one autoregressive coefficient ρ $(p = 1)$ and structural innovation $\epsilon_{t,i}$:

$$(1-\rho L)y_{t,i} \quad = \quad \epsilon_{t,i} \qquad \text{for } i = 0, 1, 2, ..., n; \; t = 1, 2, ..., T. \tag{3}$$

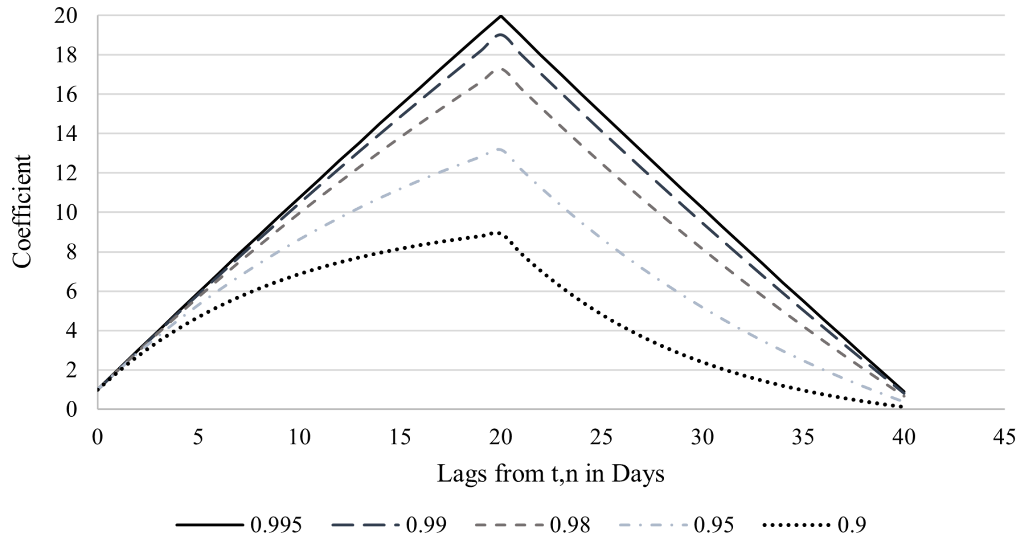
For arbitrary n, the aggregation operator is again given by $\omega(L) = 1 + L + L^2 + \cdots + L^{n-1}$, the polynomial, $\beta(L)$, is such that the product $h(L) = \beta(L)(1-\rho L)$ only contains powers of L^n. In this case, equation 2 simplifies to:

$$h(L)\bar{y}_t \quad = \quad (1 + (1+\rho)L + \cdots + \rho^{n-1}L^{2(n-1)})\epsilon_{t,i}. \tag{4}$$

Equation 4 shows coefficients of the structural innovations take the form of a non-linear polynomial and the largest lag operator on $\epsilon_{t,i}$ is $2(n-1)$.

For illustration, Figure 1 graphs the values of the coefficients on lagged daily structural innovations for alternative values of ρ for monthly average data, with $n = 21$. Daily lags from 0 to 20 represent contemporaneous innovations within the month t, and lags from 21 to 41 represent

Figure 1. Coefficients on Lagged Daily Values are Not Constant



Notes: Coefficients on lagged daily values for alternative values of ρ of an AR(1) model with temporal aggregation to monthly data, $n = 21$.

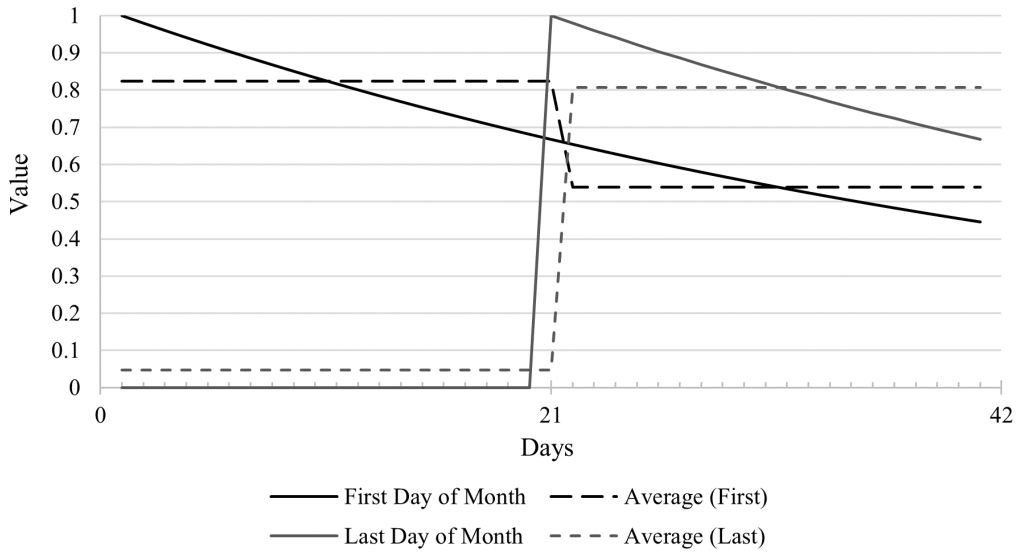
innovations from the previous month. Figure 1 clarifies that the monthly average data in time t disproportionately reflects structural innovations at the beginning of the month t and at the end of the previous month, $t-1$. In fact, the aggregated data approaches symmetric constitution of innovations from both months as $\rho \to 1$.

The implication, is that the inference on structural innovations depends on the occurrence of daily shocks within the month. For example, consider two innovations from the AR(1) model with $\rho = 0.98$, where one occurs on the first day of the month and the other occurs on the last day of the month, as shown in Figure 2. When the shock occurs on the last day of the month, the information is primarily reflected in the next month's average. Despite the daily information arriving in the previous month, the monthly average reflects less than 5 percent of the shock. Empirical use of only the aggregated data will incorrectly conclude that the innovation occurred in the succeeding month. In contrast, a shock that occurs on the first day of the month will be reflected in the current month's average.

The loss of information arises when only the monthly data is used and the sequence of daily moving average terms by approximated with an ARMA(1,1) at the monthly frequency:

$$\begin{aligned} \bar{y}_t &= \rho^n \bar{y}_{t-1} + (1 + (1+\rho)L + \cdots + \rho^{n-1} L^{2(n-1)})\epsilon_{t,i} \\ &\approx \tilde{\rho}\bar{y}_{t-1} + u_t + \tilde{\alpha} u_{t-1}. \end{aligned} \tag{5}$$

Figure 2. Information Loss Depends on the Day in the Month the Shock Occurs



Notes: Impulse response from a daily AR(1) model with $\rho - 0.98$ with monthly average data, $n = 21$.

where u_t is the estimate of the monthly structural innovation. The estimation of a monthly moving average coefficient $\tilde{\alpha}$ places a constant coefficient on the daily innovations from the previous month. This turns out to be a poor approximation for persistent data, as it fails to capture the non-constant weights imposed from the data transformation and results in substantial information loss. The difference between a common coefficient and the true coefficient is increasing in the degree of persistence. The misspecification implies that innovations from the aggregate model, u_t, exhibit $covar(u_t, L^j y_{t-1,n}) \neq 0 \; \forall \; j \in \{0, 1, ..., (n-1)\}$. Thus, predictability of the monthly structural innovations are expected using the daily information from the previous month for all $\rho > 0$.

2.2 Selective Sampling

An alternative data sampling strategy is selective sampling (point-in-time sampling). For example, time series of end-of-period data, $y_{t,n}$, are the convention for stock variables (rather than flow variables), for calculating returns to financial assets, and are commonly reported for interest rates and bilateral exchange rates.

Unfortunately, selective sampling is similar to temporal aggregation in that it can modify the structure of the data and result in information loss. Specifically, selective sampling, like temporal aggregation, is known to modify the structure of data for ARIMA (Wei, 1981; Weiss, 1984), and VARIMA processes (Marcellino, 1999) and can result in information loss when forecasting Kohn

(1982). Only in special cases can selective sampling avoid information loss.

Consider the case of the ARMA(p, q) model under selective sampling. The aggregation operator is now $\omega(L) = 1$. Then, introduce a polynomial, $\beta(L)$, which is such that the product $h(L) = \beta(L)a(L)$ only contains powers of L^n. This polynomial $\beta(L)$ always exists and degree in L is again at most equal to $p(n-1)$ (Brewer, 1973; Wei, 1981; Weiss, 1984). Following the same steps as for equation 2 we obtain

$$\begin{aligned} \omega(L)\beta(L)a(L)y_{t,i} &= (1 + \cdots + \rho^{n-1}L^{p(n-1)})(1 + \cdots + \alpha_q L^q)\epsilon_{t,i} \\ h(L)y_{t,n} &= (1 + (\rho_1 + \alpha_1)L + \cdots + \alpha_q \rho_p^{n-1} L^{p(n-1)+q})\epsilon_{t,i}. \end{aligned} \tag{6}$$

Thus, $y_{t,n}$ is said to be approximated by an ARMA(p, r) process where $r = NLI(p - (q-p)/n)$ (see, Wei, 1981; Weiss, 1984). Equation 6 shows that the coefficients of the structural innovations again take the form of a non-linear polynomial.

Thus, selective sampling still exhibits temporal bias for ARMA(p, q) representations with, $(p-1)(n-1) + q > n-1$. For example, for an AR(p) or MA(q) model, temporal bias is only present when $p \geq 2$ or $q > n-1$, respectively. Thus, selective sampling can reduce temporal bias, but can only eliminate temporal bias if the daily DGP exhibits $p \leq 1$ and low values of q relative to n.

2.3 Testing For Temporal Aggregation Bias

A method is now proposed to quantify the information loss from temporal aggregation using direct forecasts constructed from lagged daily observations. As shown in equation 2, the $covar(u_t, L^j y_{t-1,n}) \neq 0 \; \forall \; j \in \{0, 1, ..., p(n-1)+q\}$ such that a direct forecast can be constructed using daily values from the previous periods, which contains all relevant information for future period average observations. Thus, it is proposed that this information loss be quantified by constructing a direct forecast of the structural innovations on $J = p(n-1) + q$ lagged differences between the daily and monthly average observations, for stationary, defined as $d_{t,i} = y_{t,i} - \bar{y}_t$. The test takes the following form:

$$u_t = \sum_{j=0}^{J} \beta_j d_{t-1,n-j} + \nu_t. \tag{7}$$

The test quantifies the extent to which structural innovations constructed from temporally aggregated data, u_t, are predictable to economic agents by the end of the previous month, $E_{t,n}[u_{t+1}] = 0$. Note that this differs from the expectations of the econometrician who has chosen to limit information to averaged data, $\bar{E}_t[u_{t+1}] = 0$. Rejection of the latter assumption is associated with model misspecification, but is limited to the information set of the averaged data.

The test is a one-step-ahead direct forecast and has a MIxed DAta Sampling (MIDAS) specification (see for example, Ghysels et al., 2007; Andreou et al., 2010; Foroni and Marcellino, 2016; Ghysels, 2016; Andreou et al., 2010, among others). Specifically, this is an unrestricted MIDAS (UMIDAS) regression, and for small J, the parameters can be efficiently estimated with ordinary least squares (OLS) (Foroni et al., 2015, 2019). A simulation analysis of different MIDAS estimators and power analysis is provided in the next section.

With observable daily data, the test can be parameterized by first determining the data generating process of the daily data. For example, the best fit of an ARMA(p, q) model on the daily data can be used to determine p, and q, and then J is given by the sufficient conditions. When daily data is not available, lag selection can follow standard practice and use hypothesis testing or by estimating restricted MIDAS parameter profiles.

3 Quantifying Predictability

Simulation experiments are now examined to quantify the information loss from temporal aggregation.

3.1 ARMA

Daily data, $y_{t,i}$ is constructed assuming the AR(1) model of equation 13, $\epsilon_{t,i} \sim N(0, 1)$, and aggregated, $\bar{y}_t$ to weekly, monthly, or quarterly frequency, with $n = 5$, 21, or 62, respectively. Then, the ARMA(1,1) model is estimated using aggregated data, equation 5, and equation 7 is estimated on the structural innovations u_t. Simulations use 40 years worth of daily data, consistent with applications where daily data has been available since the early-1980s, and, in addition, burn the first 500 daily observations. These baseline simulations are consistent with the applications in section 4.

Table 1 reports the mean and standard deviation of the adjusted R^2 from equation 7 using 5000 simulations. The innovations are expected, rejecting $E_{t,n}[u_{t+1}] = 0$. The ability to predict shocks increases in the persistence of the daily data. When $\rho > 0.95$, daily information available in the previous period can explain over one third of the structural innovations for monthly and quarterly data. For highly persistent daily data, $\rho > 0.995$, over 37 percent of the structural innovations are

Table 1. Structural Innovations are Highly Predictable When the Data is Temporally Aggregated

Agg. Model	ARMA(1,1)			ARMA(1,0)		
ρ / Frequency	Weekly	Monthly	Quarterly	Weekly	Monthly	Quarterly
1.00	0.273	0.388	0.410	0.353	0.463	0.483
	(0.017)	(0.035)	(0.079)	(0.016)	(0.032)	(0.070)
0.995	0.271	0.375	0.370	0.349	0.444	0.429
	(0.017)	(0.035)	(0.081)	(0.016)	(0.032)	(0.074)
0.99	0.268	0.362	0.333	0.344	0.426	0.379
	(0.017)	(0.035)	(0.082)	(0.016)	(0.033)	(0.078)
0.95	0.249	0.271	0.145	0.311	0.298	0.149
	(0.016)	(0.034)	(0.090)	(0.016)	(0.034)	(0.090)
0.90	0.225	0.182	0.068	0.271	0.190	0.069
	(0.016)	(0.032)	(0.090)	(0.016)	(0.032)	(0.090)
0.75	0.151	0.061	0.022	0.166	0.062	0.022
	(0.014)	(0.024)	(0.089)	(0.015)	(0.024)	(0.089)
0.50	0.056	0.016	0.007	0.057	0.016	0.007
	(0.010)	(0.017)	(0.089)	(0.010)	(0.017)	(0.089)
0.25	0.011	0.003	0.003	0.011	0.003	0.003
	(0.005)	(0.014)	(0.089)	(0.005)	(0.014)	(0.089)
0.00	0.000	0.000	0.002	0.000	0.000	0.002
	(0.001)	(0.013)	(0.088)	(0.001)	(0.013)	(0.089)

Notes: Mean R^2-adjusted values from 5000 Monte Carlo simulations of equation 7 using UMIDAS estimated with OLS and $J = n-1$. 40 years worth of daily data simulated with an AR(1) with alternative autocorrelation coefficients ρ. "Agg. Model" refers to the model, equation 5, estimated on aggregated data to obtain the residuals, u_t. Standard deviations in parentheses.

expected for monthly and quarterly data, respectively.

Moreover, the test correctly reflects that the shocks cannot be predicted in the previous month when $\rho = 0$. This arises from the use of the R^2-adjusted measure, which is suitable since the test has a UMIDAS form. Thus, the empirical evidence indicates that there is no outfitting.[2]

The last three columns consider the case where the moving average term is omitted, and the aggregate model is estimated with an AR(1). When the moving average term is ignored, the test predicts slightly more of the structural shocks. This quantifies that not approximating the high-frequency moving average terms with an aggregate moving average term slightly exacerbates predictability. However, the primary driver of predictability is the mechanical loss of information introduced by aggregation.

Now consider the information loss from selective sampling and temporal aggregation, and we also consider when the daily data is generated with an AR(2) model. For selective sampling, the aggregate model is estimated with end-of-month data. The AR(2) model is assumed to take the

[2]Note that since the form of the error is known, approximations using restricted parameter profiles are not needed. Use of restricted parameter profiles is only found to reduce forecast precision and are not reported for brevity.

following form,

$$(1 - \rho L - \gamma L^2) y_{t,i} \quad = \quad \epsilon_{t,i} \quad \text{for } i = 0, 1, 2, ..., n; \; t = 0, 1, 2, ..., T. \qquad (8)$$

where γ is the second autoregressive coefficient. We consider two different daily AR(2) data generating processes (DGPs), a non-persistent case with $\gamma = (1 - \rho)0.5$, and a persistent case with $\gamma = (1 - \rho)0.99$. For illustration, the daily data aggregated to the monthly frequency in the simulations. The low-frequency model used to obtain the shocks, u_t, is assumed to use the aggregate approximation and includes the correct number of lower frequency moving average terms.

Table 2. Even Selectively Sampled Data can Result in Highly Predictable Structural Innovations

ρ / Sampling	AR(1)		AR(2)		Persistent AR(2)	
	EoM	Average	EoM	Average	EoM	Average
1.00	0.000	0.388	0.000	0.740	0.000	0.740
	(0.013)	(0.035)	(0.019)	(0.019)	(0.019)	(0.019)
0.995	0.000	0.375	0.000	0.708	0.000	0.739
	(0.013)	(0.035)	(0.019)	(0.021)	(0.019)	(0.019)
0.99	0.000	0.362	-0.002	0.677	0.000	0.738
	(0.013)	(0.035)	(0.019)	(0.023)	(0.019)	(0.019)
0.95	-0.001	0.271	-0.003	0.486	0.000	0.730
	(0.013)	(0.034)	(0.019)	(0.030)	(0.019)	(0.020)
0.90	-0.001	0.182	-0.003	0.337	0.000	0.720
	(0.013)	(0.032)	(0.018)	(0.033)	(0.019)	(0.020)
0.75	-0.001	0.061	-0.003	0.156	0.002	0.695
	(0.013)	(0.024)	(0.018)	(0.032)	(0.019)	(0.022)
0.50	-0.002	0.016	-0.003	0.071	0.013	0.665
	(0.013)	(0.017)	(0.018)	(0.028)	(0.021)	(0.024)
0.25	-0.002	0.003	-0.003	0.042	0.043	0.646
	(0.013)	(0.014)	(0.018)	(0.025)	(0.026)	(0.025)
0.00	-0.002	0.000	-0.004	0.028	0.390	0.634
	(0.013)	(0.013)	(0.019)	(0.024)	(0.036)	(0.025)

Notes: Mean R^2-adjusted values from 5000 Monte Carlo simulations of equation 7 using average or selective sampling. 40 years worth of daily data, using an AR(1), AR(2) with $\gamma = (1 - \rho)0.5$, or a persistent AR(2) with $\gamma = (1 - \rho)0.99$ and the first autocorrelation coefficient is given by ρ. Estimates of the low-frequency model used to obtain the shocks, u_t, is assumed to be correctly specified. Standard deviations in parentheses.

The mean and standard deviation of the adjusted R^2 using the test of equation 7 is reported in Table 2. When the daily data is an AR(1) and the data is selectively sampled, the test shows no evidence of predictability for any value of ρ, consistent with the sufficient conditions. Only when the AR(2) model has a sizable value for the second autoregressive term, γ, is there found to be predictability. For example, when $\rho = 0$ and $\gamma = 0.99$, 39 percent of the structural shocks are predictable. This is similar in magnitude to the predictability found for the AR(1) model with averaging, and shows that temporal bias could still be present with selective sampling. Moreover, when the data is temporally aggregated, up to 74 percent of the shocks are predictable when the daily DGP is an AR(2) model. Together, these finding show that up to the majority share of structural innovations can be predicted using lagged high-frequency information, and that while selective sampling can reduce predictability, it can only eliminate temporal aggregation bias in special cases.

3.2 Structural Vector Autoregressions (SVAR)

The above analysis showed that the sufficient conditions and parameterization of the tests comes from the high-frequency moving average structure derived by Brewer (1973). This was extended to vector autoregressive moving average models (VARMA) for selectively sampled and temporally aggregated data by Marcellino (1999).[3] Marcet (1991) considered the case where the variables in the VAR have different sampling methods. For temporally sampled data, the largest lag operator on the high-frequency innovations remains the same, but the test includes disaggregated observations for each variable. We use these conditions to quantify the mistiming of the structural shocks arising from the aggregation of daily data.

Consider daily data for two variables, $y_{1t,i}$ and $y_{2t,i}$, generated from a structural vector autoregressive models (SVAR). The SVAR has a recursive ordering such that the structural shock to equation 1, $\epsilon_{1t,i}$ enters equation 2 with coefficient η_2:

$$\begin{aligned} y_{1t,i} &= \rho_{11} y_{1t,i-1} + \rho_{12} y_{2t,i-1} + \epsilon_{1t,i} \\ y_{2t,i} &= \rho_{21} y_{1t,i-1} + \rho_{22} y_{2t,i-1} + \eta_2 \epsilon_{1t,i} + \epsilon_{2t,i} \end{aligned}$$

We simulate 40 years of daily data and consider three sampling cases. The first case samples both variables using monthly averages. The second case has the first variable sampled using monthly averages and the second with end-of-period sampling, while the third case reverses this sampling order. Then, the SVAR is estimated at the monthly frequency, the structural shocks are extracted, and equation 7 is estimated on the structural innovations. We consider four alternative parameterizations of the SVAR, shown in Table 3. For all cases, the first variable has a high degree of autoregressive persistence with $\rho_{11} = 0.99$ but the second variable has a low degree of autore-

[3]See also Lütkepohl (1987) and Silvestrini and Veredas (2008)

Table 3. Highly Predictable SVAR Innovations

Parameters					Ave. y_{1t}, Ave. y_{2t}		Ave. y_{1t}, Point y_{2t}		Point y_{1t}, Ave. y_{2t}	
ρ_{11}	ρ_{22}	ρ_{12}	ρ_{21}	η_2	ϵ_1	ϵ_2	ϵ_1	ϵ_2	ϵ_1	ϵ_2
0.99	0.45	0.00	0.50	0.50	0.368	0.035	0.149	0.033	-0.002	-0.001
					(0.035)	(0.023)	(0.034)	(0.024)	(0.019)	(0.019)
0.99	0.45	0.00	-0.50	0.50	0.217	0.112	0.308	0.057	-0.002	0.001
					(0.034)	(0.027)	(0.038)	(0.025)	(0.019)	(0.020)
0.99	0.45	0.05	0.10	0.10	0.459	0.010	0.412	0.006	-0.001	0.010
					(0.033)	(0.021)	(0.035)	(0.020)	(0.019)	(0.021)
0.99	0.45	0.50	-0.50	0.10	0.050	0.124	0.021	0.005	-0.002	0.030
					(0.026)	(0.032)	(0.023)	(0.020)	(0.020)	(0.024)

Notes: Mean R^2-adjusted with standard deviations in parentheses for the structural shocks, ϵ_1 and ϵ_2, from the two-variable SVAR. Daily data for $y_{1t,i}$ and $y_{2t,i}$ are sampled using monthly averages (Ave.) or end-of-month point-in-time sampling (Point). Results based on 5000 Monte Carlo simulations with 40 years of daily data.

gressive persistence, $\rho_{22} = 0.45$. This choice is illustrative, but higher autoregressive persistence in the second equation would only further increase the degree of predictability for the second shock.

Let's first consider the case where both variables are sampled using monthly averages. In the first parameterization, the second variable is influenced by both the shock and the lagged value of y_1 at the daily frequency, but not vice versa. The R^2 of the structural shock from the first equation is similar in magnitude to the AR(1) case, 0.368. The second variable has less autoregressive persistence, so the R^2 of the structural shock is only 0.035. In contrast, the second parameterization shows the case where y_{2t} responds negatively to y_{1t-1} at the daily frequency, and the R^2 for the structural shocks changed to 0.217 and 0.112, respectively. The predictability of the first shock is reduced because one-sided directional causality becomes bidirectional due to aggregation (Wei, 1982; Breitung and Swanson, 2002).

The third and fourth parameterizations introduce feedback from the second variable to the first at the daily frequency. When this feedback increases persistence in the first equation, such as with $p_{12} = 0.05$, the R^2 rises to 0.459. Conversely, when that feedback reduces persistence in the first variable, predictability declines. In the last parameterization, the second equation has a negative lagged response to y_{1t}, d $p_{21} = -0.5$. In effect, the second variable provides information from recent realizations of ϵ_{1t} that was lost in the aggregation of the first variable. This information is attributed to the second variable, lowering the predictability of the structural shock in the first equation. These results highlight the known problem of spurious relationships in multivariate systems when a right-hand-side variable contains information lost from aggregating the left-hand-side variable.

Predictability of the structural shocks persists even when the sampling methods differ between variables in the VAR. When the first variable is aggregated, and the second is point sampled, the predictability of the first variable generally decreases-except in one case. This reduction occurs because more information from recent shocks to ϵ_{1t} is preserved in the second equation and is attributed to the lagged second variable in the first equation. The second parameterization is the only case where the predictability of the first shock increases due to ρ_{21} being negative, which fails to preserve information from recent ϵ_{1t} shocks. In the final sampling case, where the first variable is point sampled, and the second is averaged but exhibits low persistence, neither shock shows predictability since the number of lagged autoregressive coefficients is less than two.

The multivariate analysis demonstrates that spurious relationships between variables and the predictability of structural shocks from temporally aggregated data are two sides of the same coin. Spurious relationships arise because a variable retains information lost through temporal aggregation of the other variable. Consequently, estimated relationships between variables may disappear when the information loss from temporal aggregation is addressed.

3.3 Power Analysis

As J becomes large, or with small sample sizes, approximations of the MIDAS structure using restricted parameter profiles may be more efficient than using UMIDAS (Foroni et al., 2015). The appropriateness of restrictions on the parameters for large J is effectively an empirical question and is now examined.

Table 4 reports the share of times that the test rejects the null hypothesis of no predictive power at the 0.05 percent level using 5000 Monte Carlo simulations of daily data generated with an AR(1). The test is estimated with OLS and $J = n - 1$ following equation 7. The model of the averaged data is correctly specified and estimated using an ARMA(1,1). Power is considered for hypothesis testing using t-tests with restricted parameters of $J = 1$, and F-tests when $J = n - 1$, and DM-tests (Diebold and Mariano, 1995) that use Newey and West (1987) standard errors and compare against the standard normal.

The results suggest that with 40 years of data, an F-test for predictability has near perfect power for $\rho \geq 0.5$ for weekly data, and $\rho \geq 0.75$ for monthly data. In contrast, power is smaller for quarterly data and requires $\rho > 0.95$. Results are very similar with only twenty years of data. When there is no predictability, $\rho = 0$, the F-test correctly reflects the 5 percent significance level.

A restricted parameter profile of $J = 1$ is sufficient to establish predictability but would not be

Table 4. Power Analysis, Weekly, Monthly, and Quarterly Frequency

	F-test ($J = n-1$)			t-test ($J = 1$)			DM ($J = n-1$)			DM ($J = 1$)		
ρ	W	M	Q	W	M	Q	W	M	Q	W	M	Q
						20 Years						
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.995	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.99	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.95	1.00	1.00	0.47	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96
0.90	1.00	1.00	0.18	1.00	1.00	0.96	1.00	1.00	1.00	1.00	1.00	0.56
0.75	1.00	0.59	0.08	1.00	0.99	0.56	1.00	1.00	1.00	1.00	0.74	0.10
0.50	1.00	0.14	0.06	1.00	0.62	0.23	1.00	0.98	1.00	1.00	0.11	0.02
0.25	0.78	0.07	0.05	0.96	0.21	0.11	0.71	0.95	1.00	0.54	0.01	0.01
0.00	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.95	1.00	0.00	0.00	0.00
						40 Years						
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.995	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.99	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.95	1.00	1.00	0.47	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96
0.90	1.00	1.00	0.19	1.00	1.00	0.96	1.00	1.00	1.00	1.00	1.00	0.56
0.75	1.00	0.95	0.08	1.00	1.00	0.57	1.00	1.00	1.00	1.00	0.99	0.10
0.50	1.00	0.29	0.06	1.00	0.86	0.24	1.00	0.99	1.00	1.00	0.31	0.02
0.25	0.98	0.08	0.05	1.00	0.34	0.10	0.97	0.97	1.00	0.93	0.02	0.00
0.00	0.06	0.05	0.05	0.04	0.05	0.05	0.04	0.95	1.00	0.00	0.00	0.00

Notes: Tested at the 0.05 significance level using 5000 Monte Carlo simulations of equation 7 using UMIDAS estimates with OLS and $J = N-1$ or $J = 1$. Daily data generated with AR(1). DM refers to the Diebold and Mariano (1995) test. The null is true for values of $\rho > 0$, and false for $\rho = 0$. W, M, and Q refer to weekly, monthly, and quarterly sampling, respectively.

expected to capture the full extent of predictability in some cases. The middle three rows of Table 4 show that $J = 1$ has more power, especially at the quarterly frequency, and when the sample is small. Interestingly, $J = 1$ is also found to explain a considerable share of the shock at the weekly and monthly frequency, see appendix. The results suggest that UMIDAS with F-tests and $J = n-1$ is sufficient for weekly and monthly data, whereas t-tests and $J = 1$ are desirable for quarterly frequency and when the sample is small.

In contrast, significance of predictability using the DM-test has less power in all cases and results in substantial type one error for monthly and quarterly data under the alternative ($\rho = 0$) with $J = n-1$. The type one error for monthly and quarterly data is consistent with existing evidence of over-rejection when the forecast and target are persistent (for example, Khalaf and Saunders, 2017; Engel and Wu, 2023). Moreover, there is substantial type two error when $\rho = 0$ and $J \neq 1$. This evidence strongly suggests that testing for predictability using Diebold and Mariano (1995) is undesirable for direct forecasts compared to standard t- and F-test, consistent with existing findings (for a summary, see Diebold, 2015).

4 Application: Shocks to the Global Market for Crude Oil

We apply the proposed test to commonly used estimates of structural shocks to the global market for crude oil (Kilian, 2009; Baumeister and Hamilton, 2019) and explore if they are unexpected. This application is warranted, since the structural vector autoregressive (SVAR) models of Kilian (2009) and Baumeister and Hamilton (2019) are very popular and employ temporally aggregated monthly data. Specifically, the U.S. refiners acquisition cost of imported crude oil used by Kilian (2009) is a monthly average, and the price of West Texas Intermediate (WTI) crude oil used by Baumeister and Hamilton (2019) is a monthly average of daily closing prices. Global crude oil production used for both studies is a monthly sum. The real economic activity index (Kilian, 2009) is a transformation of a monthly average of daily shipping rates, and world industrial production used by Baumeister and Hamilton (2019) is a monthly sum. The only variable which is not temporally aggregated is the proxy for crude oil inventories in Baumeister and Hamilton (2019) which are end-of-month stocks.

We obtain the structural shocks of Kilian (2009) by conducting an exact replication of the recursively identified structural vector autoregressive (SVAR) model.[4] The structural shocks of Baumeister and Hamilton (2019) are provided on the authors' website.[5] Both models include structural shocks to aggregate demand, crude oil supply, oil-specific demand, and Baumeister and Hamilton (2019) has an additional shock to precautionary (physical inventory) demand á la Kilian and Murphy (2014).

Daily data is available for WTI which underlies the oil-specific demand shocks and for BDI which underlies the aggregate demand shock of Kilian (2009). Let $\bar{p}_t$ be the monthly average price of WTI, and $p_{t,n}$ be the corresponding end-of-month price. The growth rate in daily versus monthly average prices is $d^p_{t,i} = ln(p_{t,i}) - ln(\bar{p}_t)$, where $ln(\cdot)$ is the natural logarithm. Similarly, let the growth rate in the daily versus monthly average Baltic Dry Index (BDI) be $d^d_{t,i} = ln(bdi_{t,i}) - ln(b\bar{d}i_t)$. Notice that nominal values are used, since the price deflator cancels out. We construct direct forecasts for

[4] See Kilian (2009) for details. Replication codes are available on the American Economic Association website, and data is provided by the U.S. Energy Information Administration (EIA) and the Index of Global Real Economic Activity from FRED, the Federal Reserve Bank of Dallas.

[5] Downloaded January 2023.

each shock series from Kilian (2009) and Baumeister and Hamilton (2019), u_t^{shk}, by estimating:

$$u_t^{shk} = \sum_{j=0}^{J} \beta_j^p d_{t-1,n-j}^p + \sum_{k=0}^{K} \beta_k^d d_{t-1,n-k}^d + \nu_t^{shk}. \quad (9)$$

Since the daily data is not available for the other shocks, lags of daily observations for both oil prices and the BDI are used. Using the daily series, AIC criteria selects a $AR(1)$ model for both the daily BDI and oil prices, so the test is implemented with $K = 20$ and $J = 21$ - the average observations within each month respectively for each series.

As recommended by the power analysis, the test for predictability is implemented using UMIDAS and F-tests with $J = n - 1$ for monthly data. However, restricted MIDAS parameter profiles are also considered using $J = 1$ with t-tests, as well as restricted MIDAS profiles using the Almon lag polynomial, and generalized exponential specifications, (see for example, Ghysels et al., 2007). Significance of predictability for restricted MIDAS specifications are tested following Diebold and Mariano (1995).

Table 5. Shocks to the Global Market for Crude Oil are Predictable

Estimate	Method	Kilian (2009)			Baumeister & Hamilton (2019)			
		Aggregate Demand	Oil Demand	Oil Supply	Aggregate Demand	Oil Demand	Oil Supply	Inventory Demand
R^2	UMIDAS	0.505	0.307	0.083	0.182	0.199	0.198	0.024
	$J = N - 1$	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.045)
	MIDAS	0.376	0.105	0.071	0.100	0.157	0.047	0.043
	Almon	(0.000)	(0.001)	(0.207)	(0.215)	(0.001)	(0.011)	(0.009)
	MIDAS	0.456	0.242	0.015	0.107	0.184	0.151	0.046
	Expon.	(0.000)	(0.000)	(0.207)	(0.222)	(0.000)	(0.018)	(0.033)
$\hat{\beta}^p$	UMIDAS	-	0.110	-0.019	0.028	0.309	-0.085	0.017
	$J = 1$	-	(0.000)	(0.010)	(0.000)	(0.000)	(0.000)	(0.083)
$\hat{\beta}^d$		0.063	-	-	0.006	-	-	-
		(0.000)	-	-	(0.073)	-	-	-
R^2		0.365	0.272	0.012	0.499	0.182	0.107	0.004
		(0.000)	(0.000)	(0.014)	(0.000)	(0.000)	(0.000)	(0.083)

Notes: Estimates of equation 9, 1983M4–2022M6, using UMIDAS, Almon lag polynomial (Almon), and generalized exponential (Expon.). Adjusted R^2 used for UMIDAS, unadjusted for MIDAS. Robust p-values in parentheses using F-tests for UMIDAS, t-tests for J=1, and Diebold and Mariano (1995) for MIDAS. β^d and β^p are the coefficient values for UMIDAS with J=1 for BDI and WTI, respectively.

The adjusted R^2 for the alternative measures of predictability using UMIDAS are presented in Table 5. With $J = 1$, the coefficient on the lagged end-of-month values, β_0, are significant at the 5 percent level in all cases, except for the inventory demand shock which also happens to be the only selected sampled data. This alone is sufficient to establish that the structural shocks reflect innovations from the previous months.

Quite a large share of the shocks are predictable. For UMIDAS, with $J = n - 1$, 50.5 percent of the aggregate demand shock of Kilian (2009) is predictable. We reject the null hypothesis of no predictability, in sample, at the 5 percent level, for every shock. Restricted MIDAS parameter profiles corroborate evidence of predictability. The results indicate that all the shocks are expected, and we reject the null hypothesis of $E_{t,n}[u_{t+1}^{shk}] = 0$. As such, the estimates of structural shocks to the global market for crude oil do not wholly represent new information.

To confirm that predictability arises entirely due to the loss of information from temporal aggregation, we also test for model misspecification. Specifically, for each shock series, we test for a moving average by estimating:

$$u_t^{shk} = \alpha \tilde{u}_{t-1}^{shk} + \tilde{u}_t^{shk}. \tag{10}$$

Table 6. Some Structural Shocks Exhibit Moving Average Processes

Estimate	Method	Kilian (2009)			Baumeister & Hamilton (2019)			
		Aggregate Demand	Oil Demand	Oil Supply	Aggregate Demand	Oil Demand	Oil Supply	Inventory Demand
$\hat{\alpha}$	MA(1)	0.006	-0.002	-0.004	-0.013	0.038	0.089	-0.089
		(0.848)	(0.932)	(0.906)	(0.566)	(0.322)	(0.016)	(0.045)

Notes: Estimates of the moving average coefficient α from equation 10, 1983M4–2022M6. Robust p-values in parentheses using t-tests.

Table 6 reports the estimates of the moving average coefficients in the structural shock series. Significant moving average coefficients are found for the oil supply and inventory demand of Baumeister and Hamilton (2019), so for these cases, we can even reject the null hypothesis of $\bar{E}_t[u_t^{shk}] = 0$. The results indicate model misspecification, potentially arising from the introduction of moving average terms through temporal aggregation.

5 Have Agents Already Responded to Predictable information?

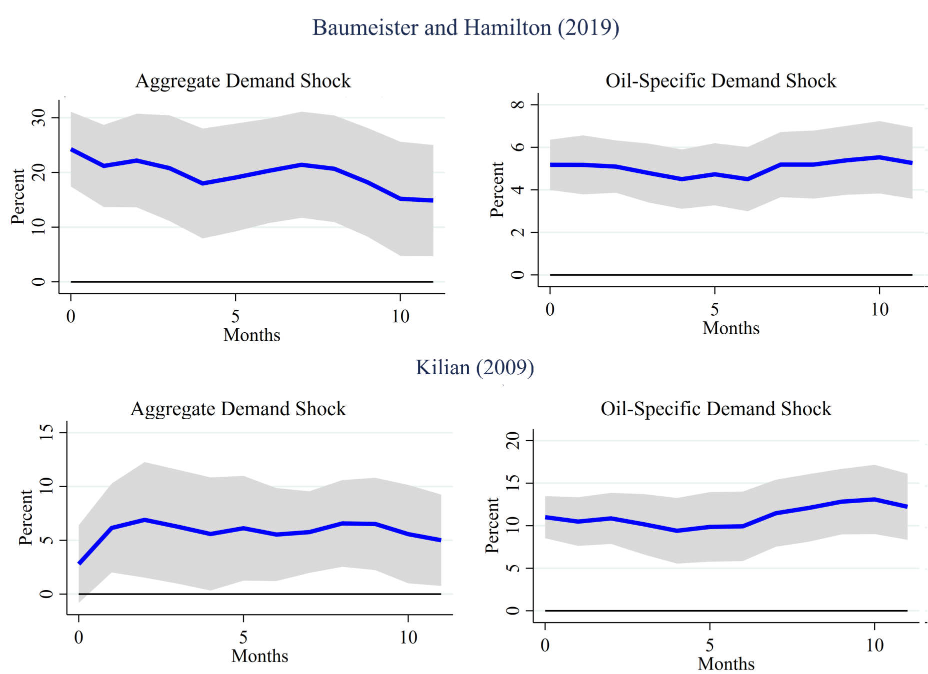
Predictable innovations suggest that the shock estimates do not wholly represent surprise innovations, as they partly reflect information from the previous period. If shocks are expected, then agents may have already partly responded to the information in the previous period, resulting in temporal aggregation bias in inference (Christiano and Eichenbaum, 1987). It is especially plausible that financial markets have already responded to the daily information from the previous month given that energy market have been shown to quickly respond to economic news (i.e. Elder et al., 2013).

Fortunately, the test of aggregation bias provides a parametric estimate to deconstruct the original shock into the new information and the expected component. Specifically, an estimate of the unexpected component is given by the estimates of the residuals, $\hat{\nu}_t^{shk}$, from equation 9. Moreover, fitted values using the estimated coefficients reflect the explained component. For the oil-specific and aggregate demand shocks, the share of the structural innovation that is expected in the previous period is given by:

$$\hat{e}_{t-1}^{shk} = \sum_{j=0}^{J} \hat{\beta}_j^p d_{t-1,n-j}^p \sum_{k=0}^{K} \hat{\beta}_k^d d_{t-1,n-j}^d. \tag{11}$$

We can test whether agents have already responded to the expected component by estimating the effects of the shocks in the previous period. We explore the dynamic effects of the expected component by constructing IRFs using local projections (Jordà, 2005). The IRFs control for the unexpected component in the current period ν_t^{shk}, and examine the response of the real price of crude oil to the expected component of the shock, $\hat{e}_t^{shk}$.

Figure 3. Markets Have Responded to Expected Component of Shocks in the Previous Month



Notes: Impulse response functions for the real price of crude oil constructed using local projections as per equation 12, 1983M6–2022M6. 90 percent confidence intervals displayed

Specifically, the IRFs for the expected component of the shock are estimated as follows:

$$y_{t+h,n} = \beta_h^u \hat{\nu}_t^{shk} + \beta_h^e \hat{e}_t^{shk} + \boldsymbol{\beta(L) Y_t} + e_{t+h}, \quad \forall h \tag{12}$$

where β_h^e is used to construct the point estimates of the IRFs, $y_{t+h,n}$ is the end-of-period real WTI price of crude oil, and Y_t controls for lagged values of the shocks and variables' in the SVAR model. Twelve lags of $\beta(L)$ are used, and confidence intervals use Newey and West (1987) standard errors.

The estimates of the IRFs for the expected component of the aggregate demand shock and the oil-specific demand shock of Kilian (2009) and Baumeister and Hamilton (2019) are presented in Figure 3. Even controlling for the effect of the unexpected shocks, the expected component of the shock has a significant effect on the real price of crude oil in the previous month. The response is very persistent and shows that the real price of crude oil had already responded to the share of the shock that was already realized in the previous period.

6 Discussion

The evidence of predictability in structural innovations presented herein is concerning and puts into question the conclusions derived using temporally aggregated data in shock identification. Use of such data, either by choice or limitation, results in information loss that implies that economic agents are likely to have already observed, and responded, to what an economist estimates as a structural innovation. This is especially plausible considering that economic agents, such those in financial markets, observe daily market data and news announcements about supply or demand changes.

If the daily data exists, e.g., daily BDI and crude oil prices, the most straightforward approach to correct for temporal aggregation bias is to avoid aggregation altogether and employ bottom-up approaches which ex-post average daily responses to the desired temporal frequency (Telser, 1967; Tiao, 1972). Alternatively, selective sampling can correct for temporal aggregation bias under the conditions set out in this paper. Finally, techniques that use high-frequency information, such as mixed frequency structural modelling, (Foroni and Marcellino, 2014, 2016; Ghysels, 2016) can also be of use. Critically, the onus should be placed on economists to test the assumption that structural innovations are unexpected.

References

Abraham, B. (1982). Temporal aggregation and time series. *International Statistical Review*, pages 285–291.

Amemiya, T. and Wu, R. Y. (1972). The effect of aggregation on prediction in the autoregressive model. *Journal of the American Statistical Association*, 67(339):628–632.

Andreou, E., Ghysels, E., and Kourtellos, A. (2010). Regression models with mixed sampling frequencies. *Journal of Econometrics*, 158(2):246–261.

Athanasopoulos, G., Hyndman, R. J., Song, H., and Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting*, 27(3):822–844.

Baumeister, C. and Hamilton, J. D. (2019). Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks. *American Economic Review*, 109(5):1873–1910.

Breitung, J. and Swanson, N. R. (2002). Temporal aggregation and spurious instantaneous causality in multiple time series models. *Journal of Time Series Analysis*, 23(6):651–665.

Brewer, K. R. (1973). Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models. *Journal of Econometrics*, 1(2):133–154.

Christiano, L. J. and Eichenbaum, M. (1987). Temporal aggregation and structural inference in macroeconomics. *Carnegie-Rochester Conference Series on Public Policy*, 26:63–130.

Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. *Journal of Business & Economic Statistics*, 33(1):1–1.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Elder, J., Miao, H., and Ramchander, S. (2013). Jumps in oil prices: the role of economic news. *The Energy Journal*, 34(3):217–237.

Engel, C. and Wu, S. P. Y. (2023). Forecasting the us dollar in the 21st century. *Journal of International Economics*, 141:103715.

Foroni, C. and Marcellino, M. (2014). Mixed-frequency structural models: Identification, estimation, and policy analysis. *Journal of Applied Econometrics*, 29(7):1118–1144.

Foroni, C. and Marcellino, M. (2016). Mixed frequency structural vector auto-regressive models. *Journal of the Royal Statistical Society: Series A*, 179(2):403–425.

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 (Statistics in Society)*, pages 57–82.

Foroni, C., Marcellino, M., and Stevanovic, D. (2019). Mixed-frequency models with moving-average components. *Journal of Applied Econometrics*, 34(5):688–706.

Georgoutsos, D. A., Kouretas, G. P., and Tserkezos, D. E. (1998). Temporal aggregation in structural var models. *Applied Stochastic Models and Data Analysis*, 14(1):19–34.

Geweke, J. (1978). Temporal aggregation in the multiple regression model. *Econometrica*, pages 643–661.

Ghysels, E. (2016). Macroeconomics and the reality of mixed frequency data. *Journal of Econometrics*, 193(2):294–314.

Ghysels, E., Sinko, A., and Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, 26(1):53–90.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, 95(1):161–182.

Khalaf, L. and Saunders, C. J. (2017). Monte carlo forecast evaluation with persistent data. *International Journal of Forecasting*, 33(1):1–10.

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99(3):1053–69.

Kilian, L. and Murphy, D. P. (2014). The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied Econometrics*, 29(3):454–478.

Kohn, R. (1982). When is an aggregate of a time series efficiently forecast by its past? *Journal of Econometrics*, 18(3):337–349.

Lütkepohl, H. (1986). Forecasting temporally aggregated vector ARMA processes. *Journal of Forecasting*, 5(2):85–95.

Lütkepohl, H. (1987). *Forecasting aggregated vector ARMA processes*, volume 284. Springer Science & Business Media.

Lütkepohl, H. (2006). Forecasting with VARMA models. *Handbook of economic forecasting*, 1:287–325.

Marcellino, M. (1999). Some consequences of temporal aggregation in empirical analysis. *Journal of Business & Economic Statistics*, 17(1):129–136.

Marcet, A. (1991). Temporal aggregation of economic time series. In Hansen, L. P. and Sargent, T. J., editors, *Rational Expectations Econometrics*, pages 237–281. Westview Press, Boulder, CO.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Rossana, R. J. and Seater, J. J. (1995). Temporal aggregation and economic time series. *Journal of Business & Economic Statistics*, 13(4):441–451.

Silvestrini, A. and Veredas, D. (2008). Temporal aggregation of univariate and multivariate time series models: a survey. *Journal of Economic Surveys*, 22(3):458–497.

Stram, D. O. and Wei, W. W. (1986). Temporal aggregation in the ARIMA process. *Journal of Time Series Analysis*, 7(4):279–292.

Teles, P. and Sousa, P. S. (2017). The effect of temporal aggregation on the estimation accuracy of time series models. *Communications in Statistics-Simulation and Computation*, 46(9):6738–6759.

Telser, L. G. (1967). Discrete samples and moving sums in stationary stochastic processes. *Journal of the American Statistical Association*, 62(318):484–499.

Tiao, G. C. (1972). Asymptotic behaviour of temporal aggregates of time series. *Biometrika*, 59(3):525–531.

Tiao, G. C. and Wei, W. S. (1976). Effect of temporal aggregation on the dynamic relationship of two time series variables. *Biometrika*, 63(3):513–523.

Wei, W. W. (1978). Some consequences of temporal aggregation in seasonal time series models. In *Seasonal analysis of economic time series*, pages 433–448. NBER.

Wei, W. W. (1981). Effect of systematic sampling on ARIMA models. *Communications in Statistics-Theory and Methods*, 10(23):2389–2398.

Wei, W. W. (1982). Comment: the effects of systematic sampling and temporal aggregation on causality–a cautionary note. *Journal of the American Statistical Association*, 77(378):316–319.

Weiss, A. A. (1984). Systematic sampling and temporal aggregation in time series models. *Journal of Econometrics*, 26(3):271–281.

Working, H. (1960). Note on the correlation of first differences of averages in a random chain. *Econometrica*, 28(4):916–918.

Zellner, A. and Montmarquette, C. (1971). A study of some aspects of temporal aggregation problems in econometric analyses. *The Review of Economics and Statistics*, pages 335–342.

A Online Appendix

A1 Additional Simulation Analysis

A1.1 Sample Size

The investigation of Section 2.1 is extended to examine the effect of sample size on predictability. Again, we assume the daily DGP is an AR(1) and the structural innovations are collected from the low-frequency model that is corrected specified for period averaging and estimated as an ARMA(1,1). The central estimates for R^2-adjusted of equation 7 are reported in Table A1.

Table A1. The Effect of Sample Size on Predictability

	Weekly			Monthly			Quarterly		
ρ / Years	10	40	500	10	40	500	20	40	500
1.00	0.272	0.273	0.274	0.383	0.388	0.390	0.402	0.410	0.417
	(0.033)	(0.017)	(0.005)	(0.078)	(0.035)	(0.010)	(0.201)	(0.079)	(0.017)
0.995	0.270	0.271	0.271	0.369	0.375	0.377	0.361	0.370	0.377
	(0.033)	(0.017)	(0.005)	(0.078)	(0.035)	(0.010)	(0.212)	(0.081)	(0.017)
0.99	0.268	0.268	0.269	0.356	0.362	0.364	0.324	0.333	0.340
	(0.033)	(0.017)	(0.005)	(0.079)	(0.035)	(0.010)	(0.223)	(0.082)	(0.017)
0.95	0.248	0.249	0.250	0.266	0.271	0.272	0.136	0.145	0.149
	(0.032)	(0.016)	(0.005)	(0.079)	(0.034)	(0.010)	(0.270)	(0.090)	(0.015)
0.90	0.224	0.225	0.225	0.178	0.182	0.183	0.062	0.068	0.069
	(0.032)	(0.016)	(0.004)	(0.078)	(0.032)	(0.009)	(0.286)	(0.090)	(0.012)
0.75	0.150	0.151	0.151	0.060	0.061	0.062	0.018	0.022	0.021
	(0.029)	(0.014)	(0.004)	(0.068)	(0.024)	(0.006)	(0.294)	(0.089)	(0.008)
0.50	0.055	0.056	0.056	0.015	0.016	0.016	0.005	0.007	0.005
	(0.020)	(0.010)	(0.003)	(0.061)	(0.017)	(0.003)	(0.299)	(0.089)	(0.007)
0.25	0.011	0.011	0.011	0.003	0.003	0.003	0.002	0.003	0.001
	(0.011)	(0.005)	(0.001)	(0.059)	(0.014)	(0.002)	(0.301)	(0.089)	(0.006)
0.00	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.000
	(0.006)	(0.001)	(0.000)	(0.058)	(0.013)	(0.001)	(0.302)	(0.088)	(0.006)

Notes: Mean adjusted R^2 values from 5000 Monte Carlo simulations of the UMIDAS equation 7 estimated with OLS. Daily data generated with AR(1) for alternative years of data, assuming 5, 21, and 63 days in a week, month and quarter, respectively. The standard deviation of the estimates is reported in parentheses.

For all sample sizes and aggregations, the use of the R^2-adjusted shows no evidence of overfitting for the case where $\rho = 0$, i.e., there is no predictability. The use of 10 or 500 years of data makes very little difference for the estimates. The results indicate that the baseline findings closely reflect the degree of predictability.

A1.2 Alternative Samplings of Daily MA(q)

Consider the case when the data generating process of the daily data is given by is a moving average model, MA(q), with common coefficient α and structural innovation $\epsilon_{t,i}$:

$$y_{t,i} \quad = \quad b(L)\epsilon_{t,i} \qquad \text{for } i = 0, 1, 2, ..., n. \tag{13}$$

where $b(L) = (1 + \alpha L + \cdots + \alpha L^q)$.

Table A2. Predictability of Innovations for Alternative Samplings: Daily MA(q) DGP

Sampling	Average					End-of-Month				
α	MA(1)	MA(10)	MA(20)	MA(30)	MA(40)	MA(1)	MA(10)	MA(20)	MA(30)	MA(40)
1.00	0.01	0.13	0.31	0.39	0.44	0.00	0.00	0.00	0.11	0.01
	(0.009)	(0.028)	(0.031)	(0.030)	(0.030)	(0.003)	(0.010)	(0.013)	(0.031)	(0.021)
0.995	0.01	0.13	0.31	0.39	0.44	0.00	0.00	0.00	0.11	0.01
	(0.009)	(0.028)	(0.031)	(0.030)	(0.030)	(0.003)	(0.010)	(0.013)	(0.031)	(0.021)
0.99	0.01	0.13	0.31	0.39	0.44	0.00	0.00	0.00	0.11	0.01
	(0.009)	(0.028)	(0.031)	(0.030)	(0.030)	(0.003)	(0.010)	(0.013)	(0.031)	(0.021)
0.95	0.01	0.13	0.31	0.38	0.44	0.00	0.00	0.00	0.11	0.01
	(0.009)	(0.028)	(0.031)	(0.030)	(0.030)	(0.003)	(0.010)	(0.013)	(0.031)	(0.021)
0.90	0.01	0.13	0.31	0.38	0.44	0.00	0.00	0.00	0.10	0.01
	(0.009)	(0.028)	(0.031)	(0.030)	(0.030)	(0.003)	(0.010)	(0.013)	(0.031)	(0.022)
0.75	0.01	0.13	0.30	0.37	0.43	0.00	0.00	0.00	0.10	0.02
	(0.009)	(0.028)	(0.031)	(0.031)	(0.030)	(0.003)	(0.010)	(0.013)	(0.030)	(0.023)
0.50	0.01	0.12	0.29	0.35	0.41	0.00	0.00	0.00	0.08	0.05
	(0.008)	(0.028)	(0.033)	(0.031)	(0.030)	(0.003)	(0.010)	(0.013)	(0.028)	(0.026)
0.25	0.00	0.09	0.24	0.30	0.36	0.00	0.00	0.00	0.07	0.09
	(0.006)	(0.026)	(0.035)	(0.033)	(0.033)	(0.003)	(0.010)	(0.013)	(0.027)	(0.030)
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.004)	(0.010)	(0.013)	(0.017)	(0.019)	(0.003)	(0.009)	(0.013)	(0.016)	(0.019)

Notes: Mean adjusted R^2 values from 5000 Monte Carlo simulations of the UMIDAS equation 7 estimated with OLS. Daily data generated with MA(q) assuming 21 days in a month. The standard deviation of the estimates is reported in parentheses.

The central estimates for R^2-adjusted for the test of equation 7 when the daily data is generated with an MA(q) and aggregated to the monthly frequency, $n = 21$, is reported in Table A2. Like the AR models, the results show that the structural shocks can be highly predictable when daily data is averaged. Moreover, while selective sampling can reduce predictability, it can only eliminate temporal aggregation bias when $q \leq n - 1$.