Granger Causality Tests with Mixed Data Frequencies

Eric Ghysels* **Rossen Valkanov†**

First Draft: July 2008

This Draft: February 18, 2009

PRELIMINARY AND INCOMPLETE

ABSTRACT

It is well known that temporal aggregation has adverse effects on Granger causality tests. Time series are often sampled at different frequencies. This is typically ignored, as data are aggregated to the common lowest frequency. The paper shows that there are unexplored advantages to test Granger causality in combining the data sampled at the different frequencies. We develop a set of Granger causality tests that take explicitly advantage of data sampled at different frequencies. Besides theoretical derivations and simulation evidence, the paper also provides an empirical application.

*Department of Finance, Kenan-Flagler Business School and Department of Economics, University of North Carolina, Gardner Hall CB 3305, Chapel Hill, NC 27599-3305, phone: (919) 966-5325, e-mail: eghysels@unc.edu.

†Rady School of Management, UCSD, Pepper Canyon Hall, 9500 Gilman Drive, MC 0093 La Jolla, CA 92093-0093, phone: (858) 534-0898, e-mail: rvalkanov@ucsd.edu.

1 Introduction

Much has been written about the spurious effects temporal aggregation may have on testing for Granger causality, see e.g. Granger (1980), Granger (1988), Lütkepohl (1993), Granger (1995), Renault, Sekkat, and Szafarz (1998), Breitung and Swanson (2002), McCrorie and Chambers (2006), among others. In this paper we deal with what might be an obvious, yet largely overlooked remedy. Time series processes are often sampled at different frequencies. Some series are available at a daily or even intra-daily frequency, other series are weekly, monthly, quarterly or even annually. Data are typically aggregated to the common lowest frequency - and therefore the availability of higher frequency data is ignored. The analysis of the paper pertains to comparing testing for Granger causality with all series aggregated to the common lowest frequency, and testing for Granger causality taking advantage of all the series sampled at whatever frequency they are available. The latter involves so called MIDAS, meaning Mi(xed) Da(ta) S(ampling), regressions introduced in Ghysels, Santa-Clara, and Valkanov (2002).[1]

There are a lot of daily and intra-daily data available, and the collection of such data is expanding as the cost of retrieval and storage have been vastly reduced over the last decade. The mixture of macroeconomic low frequency series and daily or intra-daily financial data is the most prominent example, of course. The increased availability of financial series is not only due to the reduced harvesting costs, but also due to the multiple financial innovations. For example, daily option-implied series contain market expectations of some future contingent claim outcome. Such series are starting to being used more commonly among macroeconomists. It is this type of applications that MIDAS regressions are designed for. Recent work on this topic includes, improving quarterly macro forecasts with monthly data using MIDAS regressions (see e.g. Armesto, Hernandez-Murillo, Owyang, and Piger (2008), Clements and Galvão (2008a), Clements and Galvão (2008b), Galvão (2006), Schumacher and Breitung (2008), Tay (2007)), or improving quarterly and monthly macroeconomic predictions with daily financial data (see e.g. Andreou, Ghysels, and Kourtellos (2008), Ghysels and Wright (2008), Hamilton (2006), Tay (2006)). In the context of linear time series processes it is also worth noting that the Kalman filter has be

[1] The original work on MIDAS focused on volatility predictions, see e.g. Alper, Fendoglu, and Saltoglu (2008), Chen and Ghysels (2008), Engle, Ghysels, and Sohn (2008), Forsberg and Ghysels (2006), Ghysels, Santa-Clara, and Valkanov (2005), Ghysels, Santa-Clara, and Valkanov (2006), León, Nave, and Rubio (2007), among others.

considered to handle mixed frequency data, treating as *"missing data"* to be interpolated the high frequency observations of the low frequency sampled processes, see e.g. Harvey and Pierse (1984), Harvey (1989), Bernanke, Gertler, and Watson (1997), Harvey and Chung (2000) and more recently, Aruoba, Diebold, and Scotti (2008), Bai, Ghysels, and Wright (2009) and Ghysels and Wright (2008).

We propose tests for Granger causality in mixed sampling frequency settings and derive their asymptotic properties. We also study their finite sample behavior. The paper concludes with an empirical application involving the impact of daily oil prices on monthly inflation, industrial production and quarterly GDP growth.

Section 2 starts off the paper with two motivating examples. In section 3 we study the generic setting of Granger causality with data series sampled at a low and high data frequency. In section 4 we study VAR processes - both stationary and nonstationary. In section 5 we propose various Granger causality tests and discuss their asymptotic distributions. Section 6 reports a Monte Carlo simulation study followed by section 7 containing empirical applications. Section 8 concludes the paper.

2 Motivating Examples

The purpose of this section is to provide some introductory motivating examples. To set the stage, we start with a bivariate vector autoregressive process of order one, namely:

$$\begin{aligned} y(\tau+1) &= a_{11}y(\tau) + a_{12}x(\tau) + u_y(\tau+1) \\ x(\tau+1) &= a_{21}y(\tau) + a_{22}x(\tau) + u_x(\tau+1) \end{aligned} \tag{2.1}$$

where we assume that the errors in each equation are i.i.d., in particular that $E[u_y(\tau + 1)u_x(\tau+1)] = 0$, hence so called instantaneous Granger causality is ruled out for simplicity. Suitable regularity conditions to guarantee that the VAR is weakly stationary are assumed to hold. Suppose now that the process $y(\tau)$ is only observed for τ even, whereas $x(\tau)$ is observed at all τ. We will index the time series regression that emerges for y by $t = \tau/2$ with τ even. We will also sometimes use the terms *low frequency process* for y and *high frequency* process for x. To keep track of time scales we will distinguish the process $y(\tau)$ from the process y_t. Similarly, the process $x_{t+i/2}$ is observed for all t positive or negative integers and

$i = iMod(2)$, using the convention that all odd τ are translated into fractions t - 1/2 (since τ = 2(t - 1/2)).

We start with a restricted VAR, however, where the x process is exogenous, namely:

$$\begin{aligned} y(\tau+1) &= a_{11}y(\tau) + a_{12}x(\tau) + u_y(\tau+1) \\ x(\tau+1) &= a_{22}x(\tau) + u_x(\tau+1) \end{aligned} \tag{2.2}$$

We will compare regressions involving temporally aggregated series with MIDAS regressions - to be defined shortly - when all series are stock variables for the sake of simplicity and illustrative purpose.

Let us start with the restricted system of equations (2.2). Suppose we want to test whether x Granger causes y and use the common low frequency data set - as is typically done. Namely, consider the following low frequency regression for aggregated processes:

$$y_t = \beta_1 y_{t-1} + \beta_2 x_{t-1} + \upsilon_t \tag{2.3}$$

where $\beta_1 = a_{11}^2$, $\beta_2 = = a_{12}(a_{11} + a_{22})$ and $\upsilon_t = u_t^y + a_{11}u_{t-1/2}^y + a_{12}u_{t-1/2}^x$, using again the convention that all odd τ are translated into fractions t - 1/2 for the errors. Suppose now that $\beta_2 = 0$ because $(a_{11} + a_{22}) = 0$. Then we have the spurious impression that x does not cause y.

If we take advantage of the fact that x can be sampled more frequently, we can make proper inference in this case. Namely, consider a MIDAS regression, where the low and high frequency processes are mixed. In this specific example a MIDAS regression consists of:

$$y_t = \gamma_1 y_{t-1} + \gamma_2 x_{t-1/2} + \gamma_3 x_{t-1} + \tilde{\upsilon}_t \tag{2.4}$$

where $\gamma_1 = a_{11}^2$, $\gamma_2 = a_{12}$, $\gamma_3 = a_{11}a_{12}$ and $\tilde{\upsilon}_t = u_t^y + a_{11}u_{t-1/2}^y$. Note that the orthogonality of regression errors in equation (2.2) implies that $x_{t-1/2}$ is a valid regressor, and that the MIDAS regression allows us to separately identify a_{12}. Hence, the MIDAS regression yields the right inference of one-period ahead causality patterns.

The possibility to test Granger causality from the low frequency process y to the high frequency processes x brings us to the second illustrative example. We turn now to the unconstrained bivariate system involving y and x. Suppose we are interested in testing

Granger causality via the null hypothesis $a_{21} = 0$, which is the low frequency processes causing x. Consider the following two-sided regressions:

$$y_t = \alpha_{-2} x_{t+1} + \alpha_0 x_t + \alpha_2 x_{t-1} + \upsilon_t^a \tag{2.5}$$

using aggregate processes, while for mixed data sampling we have:

$$y_t = \alpha_{-1} x_{t+1/2} + \alpha_0 x_t + \alpha_1 x_{t-1/2} + \upsilon_t^m \tag{2.6}$$

The former regression is inspired by the Sims (1972) causality regression setup. In the standard application of the Sims causality test one runs regressions with data sampled at the same frequency and one tests the significance of future regressors. Equation (2.6) is a MIDAS variation of the Sims causality test regression and will later be referred to as a *reverse MIDAS* regression as it will allow us to study indirectly projections of low frequency onto high frequency series [since a MIDAS regression involves the projection of high frequency onto low frequency].[2]

We are interested in the estimates of the lead terms as they relate to the null hypothesis. Indeed for the MIDAS regression we have that:

$$E[y_t x_{t+1/2} | x_t, x_{t-1/2}] = a_{21} Var(y_t | x_t) \tag{2.7}$$

so that the estimates of α_{-1} are directly linked to the null since $a_{21} = 0$ implies that $\alpha_{-1} = 0$. For the aggregate regression, we obtain:

$$E[y_t, x_{t+1} | x_t, x_{t-1}] = a_{21}(a_{11} + a_{22}) Var(y_t | x_t) \tag{2.8}$$

so that α_{-2} is zero whenever $a_{21} = 0$ *or* $(a_{11} + a_{22}) = 0$. Hence, the MIDAS regression ought to provide a more powerful test, whereas the information loss due to aggregation of x yields a test that may have low power or may have no power at all in certain directions, i.e. when $(a_{11} + a_{22}) = 0$.

The illustrative examples showed the potential advantages of using all the data available, despite the mixed frequencies. In the remainder of the paper we formalize the intuition provided by these two examples and provide test statistics for general VAR models.

[2]Reverse MIDAS regressions were introduced in Ghysels and Valkanov (2007).

3 The Generic Setting

We consider a K-dimensional process $\underline{x}(\tau) = (x_1(\tau), \ldots, x_K(\tau))'$, $x_i(\tau) \in L^2$, $i = 1, \ldots, K$. Suppose now that the first $K_1 < K$ elements, collected in the vector process $\underline{x}_1(\tau)$, are only observed every m periods. The remaining $K_2 = K$ - K_1 series, represented by the vector process $\underline{x}_2(\tau)$, are observed at the (high) frequency τ. As in the previous section, we will often refer to $\underline{x}_1(\tau)$ as the low frequency process, and the $\underline{x}_2(\tau)$ process as the high frequency one. For the sake of simplicity we consider the combination of two sampling frequencies. More than two sampling frequencies would amount to more complex notation, but would be conceptually similar to the analysis with a combination of two frequencies. Since the first K_1 elements are only observed every m periods, we define a time scale $t = \ldots, -1, 0, 1, \ldots,$ corresponding to all τ such that $\tau = t \times m$, where:

Assumption 3.1 (Stock Variable Skip-Sampling) *The processes $\underline{x}_{1,t} \equiv \underline{x}_1(\tau)$ is observed for $\tau = t \times m$, and the process $\underline{x}_2(\tau) \equiv x_{2,t+i/m}$ is observed for $\tau = t \times m + i$ with $i = iMod(m)$. The stacked process $x_t \equiv (\underline{x}'_{1,t}, \underline{x}'_{2,t}) \equiv \underline{x}(\tau)$ is observed for $\tau = t \times m$.*

Assumption 3.1 simplifies the analysis to a stock aggregation scheme - the case of flow variables will be discussed later. The key equations to run Granger causality tests will be the following system of MIDAS regressions:

$$x_{1,t} = \underbrace{d_t^M}_{\substack{\text{determ.}\\ \text{comp.}}} + \underbrace{\sum_{i=1}^{P} \Lambda_i^L x_{1,t-i}}_{\substack{\text{lags low freq.}\\ \text{variable}}} + \underbrace{\Lambda^I x_{2,t-1/m}}_{\substack{\text{lag high freq.}\\ \text{variable}}} + \underbrace{\Lambda^F x_{2,t+1/m}}_{\substack{\text{lead high freq.}\\ \text{variable}}} + v_{1,t} \tag{3.1}$$

Note that the above equation is a VAR model for the low frequency process $x_{1,t}$ *augmented* with high frequency lead and lag observations of $x_{2,t\pm1/m}$. The Granger causality tests between $\underline{x}_1$ and $\underline{x}_2$ will translate into parametric restrictions on the matrix coefficients Λ^I and Λ^F. Equation (3.1) shares features with (a) the original MIDAS regressions of Ghysels, Santa-Clara, and Valkanov (2002), namely the presence of high frequency data $x_{2,t-1/m}$, (b) the so called reverse MIDAS regression introduced in Ghysels and Valkanov (2007), namely the presence of low frequency data future data $x_{2,t+1/m}$, and (c) the presence of lagged dependent variables in MIDAS regressions as in Clements and Galvão (2008b). Equation (3.1) combines all these features and does it in a multivariate setting as opposed to the single regression setting typically used so far. It is the purpose of this section to explain the purpose of running the above regressions for the purpose of Granger causality testing.

Note that in equation (3.1) we only consider one high frequency lead/lag augmentation. For the purpose of formulating statistical tests, we will also consider higher order high frequency lead/lag augmentation. The key issue will be a trade-off between increase in power of statistical tests due to multiple restrictions across various lead/lags versus parameter proliferation which will typically dilute power. The issue of parameter proliferation is exactly one of the key insights in the formulation of MIDAS regressions, as will be discussed later in section 5. It will suffice for the moment to introduce multi-lead/lag extensions of (3.1), namely:

$$x_{1,t} = \underbrace{d_t^M}_{\substack{\text{determ.}\\ \text{comp.}}} + \sum_{i=1}^{P} \underbrace{\Lambda_i^L x_{1,t-i}}_{\substack{\text{lags low freq.}\\ \text{variable}}} + \sum_{i=1}^{P_b} \underbrace{\Lambda_i^I x_{2,t-i/m}}_{\substack{\text{lags high freq.}\\ \text{variable}}} + \sum_{i=1}^{P_f} \underbrace{\Lambda^F x_{2,t+i/m}}_{\substack{\text{leads high freq.}\\ \text{variable}}} + v_{1,t} \tag{3.2}$$

The new time index t also defines the relevant information accounting for the Granger causality tests. In Appendix A we define explicitly various information sets, both for the underlying high frequency processes as well as the skip-sampled ones. We will only introduce informally the most relevant information sets used in the Granger causality tests. First, for all integers t we have the *aggregate* information sets $\mathcal{I}_{i,t}^A$ $(i = 1{,}2)$ which represent univariate filtrations available at time t only involving low frequency observations of either $x_{1,t}$ $(i = 1)$ or $x_{2,t}$ $(i = 2)$. in the case of $x_{2,t}$ this means dropping the high frequency data and foregoing their information content. Also related is the joint low frequency information set: $\mathcal{I}_t^A \equiv \mathcal{I}_{1,t}^A + \mathcal{I}_{2,t}^A$. Second, for all integers t $\mathcal{I}_{2,t}^{HF}$, represents the high frequency data univariate filtration of $x_{2,\tilde{t}-i/m}$ for $\tilde{t} \leq t$ and $i = iMod(m)$. Obviously $\mathcal{I}_{2,t}^A \in \mathcal{I}_{2,t}^{HF}$. Third and last but not least we have the information set $\mathcal{I}_t^M \equiv \mathcal{I}_{1,t}^A + \mathcal{I}_{2,t}^{HF}$. This is an information set that the econometrician could collect, but typically does not, containing past $x_{1,\tilde{t}}$ *and* $x_{2,\tilde{t}-i/m}$ for $\tilde{t} \leq t$ and $i = iMod(m)$.

We also need to define information sets are defined as if the entire process would be observable all the time. The purpose of our paper is precisely to link Granger causality at high frequency with Granger causality identified via mixed data sampling. As mentioned before, the formal definitions of all the information sets are skipped here, but they can be found in Appendix A. Notably, they also include information sets $\mathcal{I}_\tau(\underline{\mathrm{x}})$, $\mathcal{I}_\tau(\underline{\mathrm{x}}_1)$ and $\mathcal{I}_\tau(\underline{\mathrm{x}}_2)$ representing the natural filtrations of the K-dimensional process $\underline{\mathrm{x}}(\tau)$ and its K_1-dimensional subvector $\underline{\mathrm{x}}_1(\tau)$

and K_2-dimensional subvector $\underline{\mathrm{x}}_2(\tau)$ respectively.

We will focus exclusively on Granger causality between the $\underline{\mathrm{x}}_1(\tau)$ and $\underline{\mathrm{x}}_2(\tau)$ via observations $x_{1,t}$ and $x_{2,t}$, without looking inside the vector processes to see whether there might be causal chains between subvectors of $x_{1,t}$ and $x_{2,t}$. Not covering tests between subvectors avoids a number of complex issues related to the distinction between single- and multi-step Granger causality. As discussed elaborately by Dufour and Renault (1998) and Dufour, Pelletier, and Renault (2006), there might be indirect causal chains in systems having at least three components. By restricting our attention to *"bivariate"* systems we do not run into the issues of one- versus multi-horizon Granger causality as they are equivalent (see Proposition 2.3 of Dufour and Renault (1998)). Since we will be interested in Granger causality between subvectors $\underline{\mathrm{x}}_1(\tau)$ and $\underline{\mathrm{x}}_2(\tau)$ we need to define:

Definition 3.1 (Granger Non-causality) *The vector $\underline{x}_i(\tau)$ does not Granger cause $\underline{x}_j(\tau)$, $i,\ j\ = 1,2,\ i \neq j$, (denoted $\underline{x}_i(\tau) \nrightarrow \underline{x}_j(\tau) \mid \mathcal{I}(\underline{x})$), if:*

$$\mathbf{P}[\underline{x}_i(\tau+1)|\mathcal{I}_\tau(\underline{x}_i)] = \mathbf{P}[\underline{x}_i(\tau+1)|\mathcal{I}_\tau(\underline{x})] \quad \forall \tau > \omega$$

where $\mathbf{P}[.|.]$ is the best linear projection onto the information sets $\mathcal{I}_\tau(\underline{x}_i)$ $i\ = 1,2$ and $\mathcal{I}_\tau(\underline{x})$ defined in Appendix A.

Note that the above definition of Granger causality is in terms of information sets we cannot actually record, due to the limitation of data collection. We will only use projections involving information sets $\mathcal{I}_{i,t}^A$ i = 1,2, $\mathcal{I}_t^A$, $\mathcal{I}_{2,t}^{HF}$, or $\mathcal{I}_t^M$, to run Granger causality tests.

It is worth noting that equation (3.2) implicitly defines the following projections:

$$\mathbf{P}[x_{1,t}|\mathcal{I}_{1,t-1}^A] = d_t^M + \sum_{i=1}^{P_M} \Lambda_i^L x_{1,t-i} \tag{3.3}$$

$$\mathbf{P}[x_{1,t}|\mathcal{I}_{t-1/m}^M] = d_t^M + \sum_{i=1}^{P_M} \Lambda_i^L x_{1,t-i} + \Lambda^I x_{2,t-1/m} \tag{3.4}$$

$$\mathbf{P}[x_{1,t}|\mathcal{I}_{t+1/m}^M] = d_t^M + \sum_{i=1}^{P_M} \Lambda_i^L x_{1,t-i} + \Lambda^I x_{2,t-1/m} + \Lambda^F x_{2,t+1/m} \tag{3.5}$$

Last but not least, we need to make the following critical assumption:

Assumption 3.2 (No Information Loss) *The low frequency process* $\underline{x}_{1,t} \equiv \underline{x}_1(\tau)$ *for* $\tau = t \times m$ *can be predicted equally well with low and high frequency data information sets:*

$$\mathbf{P}\left[\underline{x}_1(\tau+m) - \mathbf{P}[\underline{x}_1(\tau+m)|\mathcal{I}_{\tau-i\times m}(\underline{x}_1), i \geq 0]|\mathcal{I}_{\tau-i}(\underline{x}_1), i \geq 0\right] = 0$$

Note that Assumption 3.2 pertains to the *univariate* properties of the low frequency time series process. It says that univariate m-steps ahead predictions for the low frequency data using the entire high frequency history or the entire low frequency history coincide. The most prominent example is that of a VAR(1) process for $\underline{x}_1(\tau)$. Other examples include seasonal processes, say the low frequency process - if it were available daily - only involves a monthly, quarterly or yearly lag. The assumption is obviously critical as it excludes many stochastic processes, such as MA processes for which we need the high frequency filtration, or AR-type processes with more than one lag - not a multiple of m. It is important to stress, however, that the assumption only pertains to the marginal process, not the joint process for which we test Granger causality. It is worth pointing out, however, that Assumption 3.2 has implications for the joint process, namely:

Lemma 3.1 *Let Assumptions 3.2 hold. Then:*

$$\mathbf{P}\left[\underline{x}_1(\tau+m) - \mathbf{P}[\underline{x}_1(\tau+m)|\mathcal{I}_{\tau-i\times m}(\underline{x}), i \geq 0]|\mathcal{I}_{\tau-i}(\underline{x}_1), i \geq 0\right] = 0 \tag{3.6}$$

The proof of Lemma 3.1 is a rather straightforward application of the law of iterated projections, and therefore omitted.

We now state the relationship between Granger causality between high frequency data, and what we can test with mixed frequency data. Namely, the following result is obtained:

Proposition 3.1 *Let Assumptions 3.1 and 3.2 hold. Then the vector* $\underline{x}_2(\tau)$ *does not Granger cause* $\underline{x}_1(\tau)$*, implies that the parameters* Λ_i^I *in equations (3.1) and (3.2) are zero. Likewise, when the vector* $\underline{x}_1(\tau)$ *does not Granger cause* $\underline{x}_2(\tau)$*, then the parameters* Λ_i^F *in equations (3.1) and (3.2), are zero.*

The proof of Proposition 3.1 appears in Appendix B. Proposition 3.1 tells us that we can conduct inference about Granger causality relations in the high frequency joint process, despite the fact we can only partially observe at high frequency the vector process. Needless to say that this is progress with respect to the situation where we would only rely on the low frequency observations of both series.

To conclude, we turn our attention to the case of flow variables, namely we assume:

Assumption 3.3 (Flow Variable Aggregation) *The processes* $\underline{x}_{1,t} \equiv \sum_{i=0}^{m-1} \underline{x}_1(\tau - i)$ *is observed for* $\tau = t \times m$*, and the process* $\underline{x}_2(\tau) \equiv x_{2,t+i/m}$ *is observed for* $\tau = t \times m + i$ *with* $i = iMod(m)$*. The stacked process* $x_t \equiv (\underline{x}'_{1,t}, \underline{x}'_{2,t}) \equiv \underline{x}(\tau)$ *is observed for* $\tau = t \times m$*.*

It should be noted that maintaining Assumption 3.2 for flow variables is more debatable. Yet, it is easy to find a simple example that fulfills the restriction. Namely, consider a process $\underline{\mathrm{x}}_{1,t} \equiv \sum_{i=0}^{m-1} \underline{\mathrm{x}}_1(\tau - i)$, with $\underline{\mathrm{x}}_1(\tau - i) \equiv \tilde{x}_1(\tau) + \varepsilon_i$ where the latter is i.i.d. zero mean noise. In that case the flow aggregation averages out the noise. If in addition $\tilde{x}_1(\tau)$ is a 'seasonal' AR(1) with lag m we have a process that satisfies Assumptions 3.2 and 3.3. The following proposition is then proven in Appendix C:

Proposition 3.2 *Let Assumptions 3.3 and 3.2 hold. Then the vector* $\underline{x}_2(\tau)$ *does not Granger cause* $\underline{x}_1(\tau)$*, implies that the parameters* Λ_i^I *in equations (3.1) and (3.2) are zero. Likewise, when the vector* $\underline{x}_1(\tau)$ *does not Granger cause* $\underline{x}_2(\tau)$*, then the parameters* Λ_i^F *in equations (3.1) and (3.2), are zero.*

In the next section we turn to the VAR(1) model for which we can obtain stronger results.

4 The Case of Order one VAR Processes

So far we were not specific about the data generating process. In this section we focus our attention on the widely used class of VAR models, and in particular the case of an order one VAR. In particular we assume that:

Assumption 4.1 *The vector process* $\underline{x}(\tau)$*, with* $\tau \geq \omega$ *(*ω *possibly equal to* $-\infty$*) has a first order VAR representation.*

$$\underline{x}(\tau) = \underline{d}(\tau) + \Gamma \underline{x}(\tau - 1) + \underline{u}(\tau) \tag{4.1}$$

which can be written as:

$$\begin{bmatrix} \underline{x}_1(\tau) \\ \underline{x}_2(\tau) \end{bmatrix} = \begin{bmatrix} \underline{d}_1(\tau) \\ \underline{d}_2(\tau) \end{bmatrix} + \begin{bmatrix} \Gamma^{11} & \Gamma^{12} \\ \Gamma^{21} & \Gamma^{22} \end{bmatrix} \begin{bmatrix} \underline{x}_1(\tau-1) \\ \underline{x}_2(\tau-1) \end{bmatrix} + \begin{bmatrix} \underline{u}_1(\tau) \\ \underline{u}_2(\tau) \end{bmatrix} \tag{4.2}$$

where the matrices Γ^{ij} are of dimension $K_i \times K_j$, $\underline{d}(\tau)$ is a deterministic process (e.g. constant, seasonals, etc.).

In the sequel we will also assume that the error process in (4.1) is Gaussian i.i.d. with covariance matrix Ω. The distributional assumption can be relaxed, as usual, but we will assume normality to facilitate the discussion.

Note that we do not assume that the VAR process in (4.1) is covariance stationary. As noted for instance by Dufour and Renault (1998), we do not need stationarity to discuss causal relations. However, we *do* need to talk about stationary versus non-stationary VARs when we turn our attention to statistical tests. This will be deferred to section 5. We will only assume that the information regarding the $\underline{\mathrm{d}}(\tau)$ is contained in an information set $\mathcal{I}_\omega$ known at the beginning of time.

When all processes are aggregated to the common low frequency and the processes satisfy Assumption 3.1, we also have an order one VAR process, namely:

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} d_{1,t} \\ d_{2,t} \end{bmatrix} + \begin{bmatrix} \Psi^{11} & \Psi^{12} \\ \Psi^{21} & \Psi^{22} \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} \tag{4.3}$$

From equation (4.2), we can write:

$$\underline{\mathrm{x}}_1(\tau) = \underline{\mathrm{d}}_1^m(\tau-m) + (\Gamma^{11})^m \underline{\mathrm{x}}_1(\tau-m) + \sum_{j=1}^{m} (\Gamma^{11})^{j-1}\Gamma^{12}\underline{\mathrm{x}}_2(\tau-j) + \sum_{j=1}^{m} (\Gamma^{11})^{j-1}\underline{\mathrm{u}}_1(\tau-j) \tag{4.4}$$

where $\underline{\mathrm{d}}_1^m(\tau-m)$ (also denoted $\underline{\mathrm{d}}_1^m(t-m)$) is the deterministic component compounded m-steps ahead. Note that we can also rewrite equation (4.2) as a one-step ahead projection in the t time scale, augmented with lagged high frequency data as:

$$\mathbf{P}[x_{1,t+1}|\mathcal{I}^M_{(t+1)-1/m}] = \underline{\mathrm{d}}_1^m(t-m) + (\Gamma^{11})^m x_{1,t} + \Gamma^{12} \sum_{i=1}^{m} (\Gamma^{11})^{i-1} x_{2,t} \tag{4.5}$$

When we compare equation (4.2) with with equations (3.1) and (3.2), we observe mapping between the parameter λ^I (eq. (3.1)) or parameters λ_i^I (eq. (3.2)) and the high frequency VAR parameters. This will be further explored in the context of testing. Indeed, we can see the benefits and costs of using a single high frequency lag, as in equation (3.1), versus multiple lags in (3.2). For the latter, we can avoid proliferation of parameters if we either exploit the explicit link that $\lambda_i^I = (\Gamma^{11})^{i-1}\Gamma^{12}$ and therefore remain parsimonious - or else apply some of the original insights of MIDAS regressions where some generic parsimonious parametric structure is applied. Otherwise, empirical tests will most likely loose power as we increase lags in the regression equation (3.2). In contrast, in the single lag case, we do not face parameter proliferation, but we are bound to leave unexploited the possibility of increasing power through joint restrictions across lags. This matter will be discussed further in section 5.

In contrast, only using low frequency data for both x_1 and x_2 yields:

$$\mathbf{P}[x_{1,t+1}|\mathcal{I}_t^A(x)] = \underline{\mathrm{d}}_1^m(t-m) + \Theta_m^{11}x_{1,t} + \Theta_m^{12}x_{2,t} \tag{4.6}$$

where $\Theta_m^{11} = \Theta_{m-1}^{11}\Gamma^{11} + \Theta_{m-1}^{12}\Gamma^{21}$, $\Theta_m^{12} = \Theta_{m-1}^{11}\Gamma^{12} + \Theta_{m-1}^{12}\Gamma^{22}$, with $\Theta_1^{1i} = \Gamma^{1i}$, $i = 1, 2$. We note again the problems with disentangling the parameters of the high frequency VAR process, as illustrated in section 2.

Next, we turn to the reverse MIDAS formulation, namely suppose we want to compare $\mathbf{P}[x_{1,t+1}|\mathcal{I}_{(t+1)+(m-1)/m}^M]$ with $\mathbf{P}[x_{1,t+1}|\mathcal{I}_{(t+1)-1/m}^M]$. We can build on the projection appearing in (4.5) and characterize the incremental effect of adding the lead term in x_2 as it appears in equation (3.1). Using the standard results of the Frisch and Waugh (1933) theorem, we know that the coefficient of the incremental effect is zero if the following covariance is zero:

$$E\{[\sum_{i=1}^{m}(\Gamma^{11})^{i-1}\underline{\mathrm{u}}_1(\tau-i-1)][\underline{\mathrm{u}}_2(\tau) + \Gamma^{21}\sum_{i=1}^{m}(\Gamma^{11})^{i-1}\underline{\mathrm{u}}_1(\tau-i-1)] \tag{4.7}$$

The above equation implies that $\mathbf{P}[x_{1,t+1}|\mathcal{I}_{(t+1)+(m-1)/m}^M] = \mathbf{P}[x_{1,t+1}|\mathcal{I}_{(t+1)-1/m}^M]$ when Γ^{21} =0. The latter turns out to be the condition for $\underline{\mathrm{x}}_1(\tau) \nrightarrow \underline{\mathrm{x}}_2(\tau) \mid \mathcal{I}(\underline{\mathrm{x}})$. As expected from Proposition 3.1, the reverse MIDAS allows us to test whether the low frequency process x_1 does not Granger cause the high frequency process x_2. Note that having multiple leads, as in equation (3.2) results in the same parameter proliferation issues discussed with respect to equation (4.2), and therefore omitted here.

Finally, it should also be noted that replacing Assumption 3.1 by 3.3 results in similar findings, albeit algebraically more involved in terms of parameter restrictions.

5 Statistical Tests

INCOMPLETE

6 Monte Carlo Study

We conduct a Monte Carlo investigation of the Granger causality tests discussed in the previous section. We start with a parsimoneously parametrized VAR(1) model for the purpose of simulation, namely:

$$\underline{x}(\tau) = \Gamma \underline{x}_{\tau-1} + \underline{u}(\tau) \tag{6.1}$$

The matrix Γ is specified as:

$$\Gamma_1 \; = \rho \times \begin{bmatrix} 1 & \delta_l \\ \delta_h & 1 \end{bmatrix}$$

such that a single parameter ρ determines the persistence of both series, whereas δ_l and δ_h capture the dependence between the two series. The bivariate random vector $\underline{u}(\tau)$ is drawn from $N(\underline{0}, I)$, where $\underline{0}$ is a bivariate zero-vector, and I is the identity matrix of dimension 2. The data is generated for $\rho = \{0.10, 0.50, 0.90, 0.95\}$ and $\delta_h = \delta_l = \{0, -0.5, -1.5, -3.5\}$. The series are simulated for $m \times T$ observations, where $m = \{5, 10, 20, 60, 120, 250\}$, and $T = \{500, 1000\}$. We choose m to reflect empirical work. For instance, if we have daily and weekly data, then $m = 5$. If we have daily and monthly data, $m = 20$, and for daily and annual data, $m = 250$. For each of the parameters ρ, δ_l, m, and T, we simulate the series 1,000 times, where the high frequency index $\tau = 1, 2, ..., m \times T$.

We use equation (3.1) to test Granger causality and we are interested in the null hypotheses: $\Lambda_i^I = 0$, for $i = 1, \ldots m-1$. This hypothesis pertains to x_2 not Granger causing x_1, i.e. Γ^{12} is zero, or $\underline{x}_2(\tau) \nrightarrow \underline{x}_1(\tau) \mid \mathcal{I}(\underline{x})$. Conversely, we also want to test for the same equation (3.1) whether $\underline{x}_1(\tau) \nrightarrow \underline{x}_2(\tau) \mid \mathcal{I}(\underline{x})$, and this via the null hypothesis that: $\Lambda_i^F = 0$, for $i = 1, \ldots m-1$.

INCOMPLETE

7 Empirical Illustration

We turn our attention now to an empirical illustration. We examine the relation between oil prices and inflation. The latter is available at a montly frequency or lower. Oil prices can be retrieved at a daily frequency. In particular, we consider data sets that span from January 1985 - October 2008 for monthly inflation, and brent oil price series at a weekly frequency. The data sources are Thomson-Reuters Datastream for the oil price series, whereas CPI NSA series are from the Fame International Database. With these data we run regression equation (3.1) with two data frequency configurations. In Table 1 we consider a combination of quarterly inflation and monthly oil prices, whereas in Table 2 considers a combination of monthly inflation and weekly oil prices. In both Tables we report what would be standard Sims-type causality tests with leads and lags for oil prices at the same frequency as inflation - hence respectively quarterly (Table 1) and monthly (Table 2). In both cases we see clearly differences in Granger causality findings.

INCOMPLETE

8 Conclusions

INCOMPLETE

Technical Appendix

A Information sets

We consider a K-dimensional process $\underline{\mathrm{x}}(\tau) = (x_1(\tau), \ldots, x_K(\tau))'$, $x_i(\tau) \in L^2$, $i = 1, \ldots, K$. The first $K_1 < K$ elements, collected in the vector process $\underline{\mathrm{x}}_1(\tau)$, are only observed every m periods. The remaining $K_2 = K$ - K_1 series, represented by the vector process $\underline{\mathrm{x}}_2(\tau)$ are observed at the (high) frequency τ.

We divide this section into three subsections dealing with (1) the high frequency process, (2) the mixed frequency stock variable and (3) mixed frequency flow variable case.

A.1 The high frequency process

For the high frequency process we have the following information accounting:

Definition A.1 *The information set $\mathcal{I}_\omega$ contains the purely deterministic components known throughout. Moreover, for integers $\tau \in \mathbb{Z}$ the following holds:*

- *The vector processes $\underline{x}(\tau)$, $\underline{x}_1(\tau)$ and $\underline{x}_2(\tau)$ are associated with information sets $\mathcal{I}_\tau(\underline{x}) \equiv \{\underline{x}(\tilde{\tau})|\tilde{\tau} \leq \tau\}$ $+ \mathcal{I}_\omega$, $\mathcal{I}_\tau(\underline{x}_1) \equiv \{\underline{x}_1(\tilde{\tau})|\tilde{\tau} \leq \tau\}$ $+ \mathcal{I}_\omega$, and $\mathcal{I}_\tau(\underline{x}_2) \equiv \{\underline{x}_2(\tilde{\tau})|\tilde{\tau} \leq \tau\}$ $+ \mathcal{I}_\omega$, each nondecreasing sequences of closed Hilbert subspaces.*
- *The information set $\mathcal{I}_\tau(\underline{x}) \equiv \mathcal{I}_\tau(\underline{x}_1)$ $+ \mathcal{I}_\tau(\underline{x}_2)$ represents information available at time τ.*

A.2 Mixed frequency sampling with stock variables

Let Assumption 3.1 hold and the first K_1 elements are only observed every m periods, for which we define a time scale $t = \ldots, -1, 0, 1, \ldots$, corresponding to all τ such that $\tau = t \times m$, where:

Definition A.2 *The process $\underline{x}_{1,t} \equiv \underline{x}_1(\tau)$ for $\tau = t \times m$, and $\underline{x}_2(\tau) \equiv x_{2,t+i/m}$ is observed for $\tau = t \times m + i$ with $i = iMod(m)$. The process $x_t \equiv (\underline{x}'_{1,t}, \underline{x}'_{2,t})$. For all integers t the following information sets are defined:*

- *A nondecreasing sequence of closed Hilbert subspaces aggregate information sets $\mathcal{I}_t^A$, $\mathcal{I}_{i,t}^A$ (i = 1,2):*
  - $\mathcal{I}_{i,t}^A \equiv \{\underline{x}_{i,\tilde{t}}|\underline{x}_{i,\tilde{t}} \equiv \underline{x}_i(\tau), \tau = \tilde{t} \times m, \tilde{t} \leq t\}$ $+ \mathcal{I}_\omega$, *i = 1, 2.*
  - $\mathcal{I}_t^A, \equiv \mathcal{I}_{1,t}^A$ $+ \mathcal{I}_{2,t}^A$
- *A nondecreasing sequence of closed Hilbert subspaces information sets $\mathcal{I}_{2,t}^{HF}$, which represents the univariate filtration of $\mathcal{I}_{2,t}^{HF}$ $\equiv \{x_{2,\tilde{t}-i/m}|\underline{x}_{2,\tilde{t}-i/m} \equiv \underline{x}_i(\tau), \tau = \tilde{t} \times m + i, \tilde{t} \leq t, i = iMod(m)\}$ $+ \mathcal{I}_\omega$*
- $\mathcal{I}_t^M \equiv \mathcal{I}_{1,t}^A$ $+ \mathcal{I}_{2,t}^{HF}$.

A.3 Mixed frequency sampling with flow variables

We replace Assumption 3.1 by Assumption 3.3. Again, the first K_1 elements are only observed every m periods, for which we define a time scale $t = \ldots, -1, 0, 1, \ldots$, corresponding to all τ such that $\tau = t \times m$, where:

Definition A.3 *The process $\underline{x}_{1,t} = \sum_{i=0}^{m-1} \underline{x}_1(\tau - i) for\ \tau = t \times m.$ is a flow variable, whereas $\underline{x}_2(\tau) \equiv x_{2,t+i/m}$ is observed for $\tau = t \times m + i$ with $i = iMod(m)$. For all integers t the following information sets are defined:*

- *A nondecreasing sequence of closed Hilbert subspaces aggregate information sets $\mathcal{I}_t^A$, $\mathcal{I}_{i,t}^A$ (i = 1,2):*
  - $\mathcal{I}_{i,t}^A \equiv \{\underline{x}_{i,\tilde{t}} | \underline{x}_{i,\tilde{t}} \equiv \sum_{i=0}^{m-1} \underline{x}_i(\tau - i), \tau = \tilde{t} \times m, \tilde{t} \leq t\} + \mathcal{I}_\omega,\ i = 1,\ 2.$
  - $\mathcal{I}_t^A, \equiv \mathcal{I}_{1,t}^A + \mathcal{I}_{2,t}^A$
- *A nondecreasing sequence of closed Hilbert subspaces information sets $\mathcal{I}_{2,t}^{HF}$, which represents the univariate filtration of $\mathcal{I}_{2,t}^{HF} \equiv \{x_{2,\tilde{t}-i/m} | \underline{x}_{2,\tilde{t}-i/m} \equiv \underline{x}_i(\tau), \tau = \tilde{t} \times m + i, \tilde{t} \leq t, i = iMod(m)\} + \mathcal{I}_\omega$*
- $\mathcal{I}_t^M \equiv \mathcal{I}_{1,t}^A + \mathcal{I}_{2,t}^{HF}$.

B Proof of Proposition 3.1

We operate under the information accounting appearing in subsection A.2.

Let us first consider the case of $\underline{x}_2(\tau)$ not Granger causing $\underline{x}_1(\tau)$, i.e. $\underline{x}_2(\tau) \nrightarrow \underline{x}_1(\tau) \mid \mathcal{I}(\underline{x})$. We want to show that $\Lambda^I = 0$ in equation (3.1), $\Lambda_i^I = 0\ \forall\ i$ in equation (3.2), or equivalently $\mathbf{P}[x_{1,t} | \mathcal{I}_{1,t-1}^A] = \mathbf{P}[x_{1,t} | \mathcal{I}_{t-1/m}^M]$.

Using Assumption 3.1 and the property of iterated projections we know that:

$$\mathbf{P}[x_{1,t} | \mathcal{I}_{t-1/m}^M] = \mathbf{P}\left[\mathbf{P}[\underline{x}_1(tm) | \mathcal{I}_{(tm-1)}(\underline{x})] | \mathcal{I}_{t-1/m}^M\right] \tag{B.1}$$

$$= \mathbf{P}\left[\mathbf{P}[\underline{x}_1(tm) | \mathcal{I}_{(tm-1)}(\underline{x}_1)] | \mathcal{I}_{t-1/m}^M\right] \tag{B.2}$$

where we use in equation (B.2) the fact $\underline{x}_2(\tau) \nrightarrow \underline{x}_1(\tau) \mid \mathcal{I}(\underline{x})$. From Proposition 2.3 of Dufour and Renault (1998) (so called Exhaustivity Condition for Noncausality at all Horizons) we know that one-horizon Granger noncausality also implies multi-horizon noncausality. Hence, equation (B.2) also implies:

$$\mathbf{P}[x_{1,t} | \mathcal{I}_{t-1/m}^M] = \mathbf{P}\left[\mathbf{P}[\underline{x}_1(tm) | \mathcal{I}_{(tm-m)}(\underline{x}_1)] | \mathcal{I}_{t-1/m}^M\right] \tag{B.3}$$

Note that:

$$\mathbf{P}[\underline{x}_1(tm) | \mathcal{I}_{(tm-m)}(\underline{x}_1)] = \mathbf{P}[\underline{x}_1(tm) | \mathcal{I}_{1,t-1}^A] + \mathbf{P}[\underline{x}_1(tm) - \mathbf{P}[\underline{x}_1(tm) | \mathcal{I}_{1,t-1}^A] | \mathcal{I}_{(tm-m)}(\underline{x}_1)] \tag{B.4}$$

Hence, combining (B.3) and (B.4) yields:

$$\mathbf{P}[x_{1,t} | \mathcal{I}_{t-1/m}^M] = \mathbf{P}[\underline{x}_1(tm) | \mathcal{I}_{1,t-1}^A] + \mathbf{P}\left[\underline{x}_1(tm) - \mathbf{P}[\underline{x}_1(tm) | \mathcal{I}_{1,t-1}^A] | \mathcal{I}_{1,t-1}^{HF}\right] \tag{B.5}$$

Finally, using Assumption 3.2, we know we can condition only on the low frequency information set, so that the last term on the right hand side disappears. Therefore:

$$\begin{aligned}\mathbf{P}[x_{1,t}|\mathcal{I}^{M}_{t-1/m}] &= \mathbf{P}\left[\mathbf{P}[x_{1,t}|\mathcal{I}^{A}_{1,t-1}]|\mathcal{I}^{M}_{t-1/m}\right] \\ &= \mathbf{P}[x_{1,t}|\mathcal{I}^{A}_{1,t-1}] \end{aligned} \tag{B.6}$$

which means $\Lambda^I = 0$ in equation (3.1) and $\Lambda^I_i = 0 \; \forall \; i$ in equation (3.2).

Next, we consider the case of $\underline{x}_1(\tau)$ not Granger causing $\underline{x}_2(\tau)$, i.e. $\underline{x}_1(\tau) \nrightarrow \underline{x}_2(\tau) \mid \mathcal{I}(\underline{x})$. We want to show that $\Lambda^F = 0$ in equation (3.1), $\Lambda^F_i = 0 \; \forall \; i$ in equation (3.2) or equivalently $\mathbf{P}[x_{1,t}|\mathcal{I}^{A}_{1,t-1} + \mathcal{I}^{HF}_{2,t-1/m} + x_{2,t+1/m}]$ $= \mathbf{P}[x_{1,t}|\mathcal{I}^{A}_{1,t-1} + \mathcal{I}^{HF}_{2,t-1/m}]$.

Using Theorem 2 of Sims (1972), we know that $\underline{x}_1(\tau) \nrightarrow \underline{x}_2(\tau) \mid \mathcal{I}(\underline{x})$ implies that:

$$\mathbf{P}[\underline{x}_1(\tau)|\mathcal{I}_{(\tau-1)}(\underline{x}) + \mathcal{I}_{(\tau+1)}(\underline{x}_2)] = \mathbf{P}[\underline{x}_1(\tau)|\mathcal{I}_{(\tau-1)}(\underline{x})] \tag{B.7}$$

Projecting both sides of the above equation onto $\mathcal{I}_{(\tau-m)}(\underline{x}_1) + \mathcal{I}_{(\tau+1)}(\underline{x}_2)$, we have:

$$\mathbf{P}[\underline{x}_1(\tau)|\mathcal{I}_{(\tau-m)}(\underline{x}_1) + \mathcal{I}_{(\tau+1)}(\underline{x}_2)] = \mathbf{P}[\underline{x}_1(\tau)|\mathcal{I}_{(\tau-m)}(\underline{x}_1) + \mathcal{I}_{(\tau-1)}(\underline{x}_2)] \tag{B.8}$$

Next, we consider:

$$\begin{aligned}&\mathbf{P}[x_{1,t}|\mathcal{I}^{A}_{1,t-1} + \mathcal{I}^{HF}_{2,t-1/m} + x_{2,t+1/m}] \\ &\quad = \mathbf{P}\left[\mathbf{P}[\underline{x}_1(tm)|\mathcal{I}_{(tm-m)}(\underline{x}_1) + \mathcal{I}_{(tm+1)}(\underline{x}_2)]|\mathcal{I}^{A}_{1,t-1} + \mathcal{I}^{HF}_{2,t-1/m} + x_{2,t+1/m}\right] \\ &\quad = \mathbf{P}\left[\mathbf{P}[\underline{x}_1(tm)|\mathcal{I}_{(tm-m)}(\underline{x}_1) + \mathcal{I}_{(tm-1)}(\underline{x}_2)]|\mathcal{I}^{A}_{1,t-1} + \mathcal{I}^{HF}_{2,t-1/m} + x_{2,t+1/m}\right]\end{aligned} \tag{B.9}$$

where the latter follows from equation (B.8), or equivalently $\underline{x}_1(\tau) \nrightarrow \underline{x}_2(\tau) \mid \mathcal{I}(\underline{x})$. We can rewrite (B.9) as follows:

$$\begin{aligned}&\mathbf{P}\left[\mathbf{P}[\underline{x}_1(tm)|\mathcal{I}_{(tm-m)}(\underline{x}_1) + \mathcal{I}_{(tm-1)}(\underline{x}_2)]|\mathcal{I}^{A}_{1,t-1} + \mathcal{I}^{HF}_{2,t-1/m} + x_{2,t+1/m}\right] \\ &\quad = \mathbf{P}\left[\mathbf{P}[\underline{x}_1(tm)|\mathcal{I}^{A}_{1,t-1} + \mathcal{I}^{HF}_{2,t-1/m}]|\mathcal{I}^{A}_{1,t-1} + \mathcal{I}^{HF}_{2,t-1/m} + x_{2,t+1/m}\right] + \mathbf{P}\left[\varepsilon_t|\mathcal{I}^{A}_{1,t-1} + \mathcal{I}^{HF}_{2,t-1/m} + x_{2,t+1/m}\right]\end{aligned} \tag{B.10}$$

where:

$$\varepsilon_t = \mathbf{P}[\underline{x}_1(tm) - \mathbf{P}[\underline{x}_1(tm)|\mathcal{I}^{A}_{2,t-1} + \mathcal{I}^{HF}_{1,t-1/m}]|\mathcal{I}_{(tm-m)}(\underline{x}_1) + \mathcal{I}_{(tm-1)}(\underline{x}_2)] \tag{B.11}$$

Next we apply Lemma 3.1, which implies that ε_t is zero in equation (B.11) and combining equations (B.9) and (B.10) yields:

$$\mathbf{P}[x_{1,t}|\mathcal{I}^{A}_{1,t-1} + \mathcal{I}^{HF}_{2,t-1/m} + x_{2,t+1/m}] = \mathbf{P}[x_{1,t}|\mathcal{I}^{A}_{1,t-1} + \mathcal{I}^{HF}_{2,t-1/m}] \tag{B.12}$$

which implies that $\Lambda^F = 0$ in equation (3.1), $\Lambda_i^F = 0 \; \forall \; i$ in equation (3.2). □

C Proof of Proposition 3.2

We replace information accounting in subsection A.3 with that applicable to flow variables as appearing in subsection A.3.

Using Assumption 3.3, we can again Proposition 2.3 of Dufour and Renault (1998) showing that one-horizon Granger noncausality also implies multi-horizon noncausality. Hence, equation (B.3) now becomes:

$$\mathbf{P}[x_{1,t}|\mathcal{I}^M_{t-1/m}] \quad = \quad \mathbf{P}\left[\mathbf{P}[\sum_{i=1}^{m-1} \underline{\mathrm{x}}_1(tm-i)|\mathcal{I}_{(tm-m)}(\underline{\mathrm{x}}_1)]|\mathcal{I}^M_{t-1/m}\right] \tag{C.1}$$

Then with appropriate interpretations of $x_{1,t}$ and the information set $I^A_{1,t-1}$, we obtain the result in equation (B.6).

Next, we consider again the case of $\underline{\mathrm{x}}_1(\tau)$ not Granger causing $\underline{\mathrm{x}}_2(\tau)$, i.e. $\underline{\mathrm{x}}_1(\tau) \nrightarrow \underline{\mathrm{x}}_2(\tau) \mid \mathcal{I}(\underline{\mathrm{x}})$. Theorem 2 of Sims (1972) combined with equation (B.8) also implies the following for $i = 0, 1, \ldots, m-1$:

$$\mathbf{P}[\underline{\mathrm{x}}_1(\tau-i)|\mathcal{I}_{(\tau-m)}(\underline{\mathrm{x}}) + \mathcal{I}_{(\tau+1)}(\underline{\mathrm{x}}_2)] \quad = \quad \mathbf{P}[\underline{\mathrm{x}}_1(\tau-i)|\mathcal{I}_{(\tau-m)}(\underline{\mathrm{x}}) + \mathcal{I}_{(\tau-1)}(\underline{\mathrm{x}}_2)] \tag{C.2}$$

We can then take the sum of $i = 0, 1, \ldots, m-1$ and proceed to equation (B.9) and proceed with the rest of the proof using again the appropriate interpretations of $x_{1,t}$ and the information set $I^A_{1,t-1}$. □

References

Alper, C., Fendoglu, S., and Saltoglu, B. (2008), "Forecasting Stock Market Volatilities Using MIDAS Regressions: An Application to the Emerging Markets," MPRA Paper No. 7460.

Andreou, E., Ghysels, E., and Kourtellos, A. (2008), "Should Macroeconomic Forecasters look at Daily Financial Data?" Tech. rep., Discussion Paper UNC and University of Cyprus.

Armesto, M., Hernandez-Murillo, R., Owyang, M., and Piger, J. (2008), "Measuring the Information Content of the Beige Book: A Mixed Data Sampling Approach," *Journal of Money, Credit and Banking (forthcoming)*.

Aruoba, S., Diebold, F., and Scotti, C. (2008), "Real-Time Measurement of Business Conditions," Working paper, University of Pennsylvania.

Bai, J., Ghysels, E., and Wright, J. (2009), "A Daily Index of Macroeconomic Activity," Tech. rep., Work in Progress.

Bernanke, B., Gertler, M., and Watson, M. (1997), "Systematic Monetary Policy and the Effects of Oil Price Shocks," *Brookings Papers on Economic Activity*, 1, 91–157.

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

Chen, X. and Ghysels, E. (2008), "News-Good or Bad-and its Impact Over Multiple Horizons," Discussion Paper UNC.

Clements, M. and Galvão, A. (2008a), "Forecasting US output growth using Leading Indicators: An appraisal using MIDAS models," *Journal of Applied Econometrics (forthcoming)*.

— (2008b), "Macroeconomic Forecasting with Mixed Frequency Data: Forecasting US output growth," *Journal of Business and Economic Statistics*, 26, 546–554.

Dufour, J., Pelletier, D., and Renault, E. (2006), "Short run and long run causality in time series: inference," *Journal of Econometrics*, 132, 337–362.

Dufour, J. M. and Renault, E. (1998), "Short run and long run causality in time series: Theory," *Econometrica*, 66, 1099–1125.

Engle, R., Ghysels, E., and Sohn, B. (2008), "On the Economic Sources of Stock Market Volatility," Discussion Paper NYU and UNC.

Forsberg, L. and Ghysels, E. (2006), "Why do absolute returns predict volatility so well?" *Journal of Financial Econometrics*, 6, 31–67.

Frisch, R. and Waugh, F. V. (1933), "Partial Time Regressions as Compared with Individual Trends," *Econometrica*, 1, 387–401.

Galvão, A. (2006), “Changes in Predictive Ability with Mixed Frequency Data,” Discussion Paper QUeen Mary.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2002), “The MIDAS touch: Mixed data sampling regression models,” Working paper, UNC and UCLA.

— (2005), “There is a risk-return tradeoff after all,” *Journal of Financial Economics*, 76, 509–548.

— (2006), “Predicting volatility: getting the most out of return data sampled at different frequencies,” *Journal of Econometrics*, 131, 59–95.

Ghysels, E. and Valkanov, R. (2007), “Linear Time Series Processes with Mixed Data Sampling and MIDAS Regression Models,” Working paper UNC and UCSD.

Ghysels, E. and Wright, J. (2008), “Forecasting professional forecasters,” *Journal of Business and Economic Statistics (forthcoming).*

Granger, C. (1980), “Testing for Causality: A Personal Viewpoint,” *Journal of Economic Dynamics and Control*, 2, 329–352.

— (1988), “Some Recent Developments in A Concept of Causality,” *Journal of Econometrics*, 39, 199–211.

— (1995), “Causality in the Long Run,” *Econometric Theory*, 11, 530–536.

Hamilton, J. (2006), “Daily Monetary Policy Shocks and the Delayed Response of New Home Sales,” Working paper, UCSD.

Harvey, A. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

Harvey, A. and Chung, C. (2000), “Estimating the underlying change in unemployment in the UK,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 163, 303–309.

Harvey, A. C. and Pierse, R. G. (1984), “Estimating missing observations in economic time series,” *Journal of the American Statistical Association*, 79, 125–131.

León, Á., Nave, J., and Rubio, G. (2007), “The relationship between risk and expected return in Europe,” *Journal of Banking and Finance*, 31, 495–512.

Lütkepohl, H. (1993), “Testing for Causation Between Two Variables in Higher Dimensional VAR Models,” in *Studies in Applied Econometrics*, eds. Schneeweiss, H. and Zimmerman, K., Heidelberg: Springer-Verlag, p. 75.

McCrorie, J. and Chambers, M. (2006), “Granger causality and the sampling of economic processes,” *Journal of Econometrics*, 132, 311–336.

Renault, E., Sekkat, K., and Szafarz, A. (1998), "Testing for spurious causality in exchange rates," *Journal of Empirical Finance*, 5, 47–66.

Schumacher, C. and Breitung, J. (2008), "Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data," *International Journal of Forecasting*, 24, 386–398.

Sims, C. (1972), "Money, Income, and Causality," *American Economic Review*, 62, 540–552.

Tay, A. (2006), "Financial Variables as Predictors of Real Output Growth," Discussion Paper SMU.

— (2007), "Mixing Frequencies: Stock Returns as a Predictor of Real Output Growth," Discussion Paper SMU.

Table 1: Developed Countries - Quarterly Inflation and Brent Oil Price Series

The entries are point estimates and standard errors for equation (3.1), namely:

$$x_{1,t} = d_t^M + \sum_{i=1}^{P} \Lambda_i^L x_{1,t-i} + \Lambda^I x_{2,t-1/m} + \Lambda^F x_{2,t+1/m} + v_{1,t}$$

where d_t^M is a constant, using quarterly inflation for x_1 and end-of-the-month oil prices for x_2. with data from January 1985 - October 2008 for monthly inflation, and brent oil price series at a weekly frequency. The data sources are Thomson-Reuters Datastream for the oil price series, whereas CPI NSA series are from the Fame International Database. The colums $m-1$ and $m+1$ pertain to lag/leads for monthly oil prices whereas the columns $q-1$ and $q+1$ represent standard Sims-type test involving quarterly series for both inflation and oil prices.

Country	m-1	m+1	q-1	q+1
United States	.01575	.00606	-.001	.00119
	.0032	.00312	.00307	.00335
Canada	.01638	-.00359	-.00535	.00425
	.00489	.00336	.00409	.00378
France	.00083	-.00063	-.00073	.01039
	.00275	.00286	.00159	.00146
Germany	.00642	.01139	-.00291	-.00152
	.00358	.00374	.0028	.00263
Italy	.00426	.00572	-.00268	.00105
	.00239	.0022	.00191	.00262
Japan	.00477	-.00388	-.00469	.00743
	.0034	.00582	.00269	.00234
Norway	.01256	.00144	-.00616	-.00381
	.00572	.00561	.00431	.00433
U.K.	.00447	.00003	.00137	.00167
	.00429	.00413	.00289	.00373

Table 2: Developed Countries - Monthly Inflation and Brent Oil Price Series

The entries are point estimates and standard errors for equation (3.1), namely:

$$x_{1,t} = d_t^M + \sum_{i=1}^{P} \Lambda_i^L x_{1,t-i} + \Lambda^I x_{2,t-1/m} + \Lambda^F x_{2,t+1/m} + v_{1,t}$$

where d_t^M is a constant, using monthly inflation for x_1 and end-of-the-week oil prices for x_2. with data from January 1985 - October 2008 for monthly inflation, and brent oil price series at a weekly frequency. The data sources are Thomson-Reuters Datastream for the oil price series, whereas CPI NSA series are from the Fame International Database. The colums $w-1$ and $w+1$ pertain to lag/leads for weekly oil prices whereas the columns $m-1$ and $m+1$ represent standard Sims-type test involving monthly series for both inflation and oil prices.

Country	w-1	w+1	m-1	m+1
United States	-.00633	.00484	-.00052	-.00126
	.00252	.00218	.00136	.0011
Canada	.00358	.00434	.0018	-.0009
	.00591	.00308	.00185	.00181
France	-.00054	.00104	-.00203	-.00047
	.00256	.00203	.00095	.00101
Germany	.00563	.00046	.00178	.00133
	.00281	.00219	.00119	.00121
Italy	.00024	.00269	.00057	.00061
	.00203	.00126	.0009	.00087
Japan	.00199	-.00036	.00047	-.00085
	.00345	.00307	.00138	.00162
Norway	-.00358	.00057	.00233	.00027
	.00469	.00436	.00194	.00195
U.K.	-.00089	.0004	-.0022	-.00047
	.00294	.00246	.00144	.00142