The MIDAS Touch: Mixed Data Sampling Regression Models*

Eric Ghysels
Department of Economics
University of North Carolina†
and CIRANO

Pedro Santa-Clara
The Anderson School
UCLA‡

Rossen Valkanov
The Anderson School
UCLA§

First Draft: February 2002
This Draft: June 22, 2004

Abstract

We introduce Mixed Data Sampling (henceforth MIDAS) regression models. The regressions involve time series data sampled at different frequencies. Technically speaking MIDAS models specify conditional expectations as a distributed lag of regressors recorded at some higher sampling frequencies. We examine the asymptotic properties of MIDAS regression estimation and compare it with traditional distributed lag models. MIDAS regressions have wide applicability in macroeconomics and finance.

*We thank Tim Bollerslev, Mike Chernov, Rob Engle, John Geweke, David Hendry, Nour Meddahi, Eric Renault, Neil Shephard, Chuck Whiteman, Jonathan Wright as well as seminar participants at City University of Hong Kong, Emory University, the Federal Reserve Board, ITAM, Korea University, New York University, Oxford University, Tsinghua University, University of Iowa, UNC and USC. All remaining errors are our own.

†Gardner Hall CB 3305, Chapel Hill, NC 27599-3305, phone: (919) 966-5325, e-mail: eghysels@unc.edu.
‡Los Angeles, CA 90095-1481, phone: (310) 206-6077, e-mail: pedro.santa-clara@anderson.ucla.edu.
§Los Angeles, CA 90095-1481, phone: (310) 825-7246, e-mail: rossen.valkanov@anderson.ucla.edu.

1 Introduction

A typical time series regression model involves data sampled at the same frequency. The idea to construct regression models that combine data with different sampling frequencies is relatively unexplored.[1] We discuss various ways to construct such regressions. We call the regression framework a Mi(xed) Da(ta) S(ampling) regression (henceforth MIDAS regression). At a general level, the interest in MIDAS regressions addresses a situation often encountered in practice where the relevant information is high frequency data, whereas the variable of interest is sampled at a lower frequency. One example pertains to models of stock market volatility. The low frequency variable is for instance the quadratic variation or other volatility process over some long future horizon corresponding to the time to maturity of an option, whereas the high frequency data set is past market information potentially at the tick-by-tick level. Yet another MIDAS example arises in the modelling of Value-at-Risk which attempts to forecast likely future losses using quantiles of the (conditional) portfolio return distribution. The horizon of interest is usually 10 days, whereas the information is again market-driven and abundant at intra-daily sampling frequencies. The interest in MIDAS regressions can also result from limitations to data availability. For example, some macroeconomic data are sampled monthly, like price series and monetary aggregates, whereas other series are sampled quarterly or annually, typically real activity series like GDP and its components. Take for instance the relationship between inflation and growth. Instead of aggregating the inflation series to a quarterly sampling frequency to match GDP data, one can run a MIDAS regression combining monthly and quarterly data.

MIDAS involve regressors with different sampling frequencies and are therefore *not* autoregressive models, since the notion of autoregression implicitly assumes that data are sampled at the same frequency in the past. Instead MIDAS regressions share some features with distributed lag models but also have unique novel features. A stylized distributed lag model is a regression of the following type: $Y_t = \beta_0 + B(L)X_t + \varepsilon_t$, where $B(L)$ is some finite or infinite lag polynomial operator, usually parameterized by a small set of hyperparameters.[2] To introduce MIDAS regressions, suppose that Y_t is sampled at some

[1] One notable exception is a chapter in John Geweke's PhD thesis (Geweke (1975), Chap. 8), which will be further discussed in the paper, where he studied mixed temporal aggregation with heterogeneous observational frequencies.

[2] See e.g. Dhrymes (1971) and Sims (1974) for surveys on distributed lag models. Many econometrics textbooks also cover the topic, see e.g. Greene (2000) (chap. 17), Judge, Griffith, Hill, Lutkepohl, and Lee (1985) (chap. 9 - 10), Stock and Watson (2003) (chap. 13), Wooldridge (1999) (chap. 18), among others.

fixed, say annual, quarterly, monthly or daily, sampling frequency and call this the interval of reference. Moreover, let $X^{(m)}$ be sampled m times faster, so that for example with annual data and $m = 4$, $X^{(4)}$ is sampled quarterly.[3] Using this notation, we can write a simple linear MIDAS regression: $Y_t = \beta_0 + B(L^{1/m})\ X_t^{(m)} + \varepsilon_t^{(m)}$ where $B(L^{1/m}) = \sum_{j=0}^{j^{max}} B(j)L^{j/m}$ is a polynomial of length j^{max} (possibly infinite) in the $L^{1/m}$ operator, and $L^{j/m}X_t^{(m)}=X_{t-j/m}^{(m)}$. In other words, the $L^{j/m}$ operator produces the value of $X_t^{(m)}$ lagged by j/m periods. The annual/quarterly example would imply that the above equation is a projection of yearly Y_t onto quarterly data $X_t^{(m)}$ using up to j^{max} quarterly lags.[4]

There are differences and similarities between distributed lag models and MIDAS regressions. Our goal is to present a general discussion of model specification and estimation in mixed sampling frequency settings, starting with a comparison of MIDAS and distributed lag models and then proceeding with more general MIDAS models. On the surface, the econometric estimation issues appear straightforward, since MIDAS regression models involve (nonlinear) least squares or related procedures. However, when it is recognized that any sampling frequency can be mixed with any other, and that potential approximation errors may come into play, one faces some challenging econometric issues. Some of these issues are addressed, others remain open questions. For example, MIDAS regressions relate to temporal aggregation issues. The mathematical structure commonly adopted to study aggregation is one that assumes that the underlying stochastic processes evolve in continuous time and data are collected at equi-distant discrete points in time. Formulating a model in continuous time has the appeal of a priori imposing a structure on discretely observed data that is independent of the sampling interval. It is this appeal that explains the considerable literature on continuous time models, a very partial list of papers studying various aspects of such models includes Bergstrom (1990), Chambers (1991), Comte and Renault (1996), Geweke (1978), Hansen and Sargent (1983), Hansen and Sargent (1991a), Hansen and Sargent (1991b), McCrorie (2000), Phillips (1959), Phillips (1972), Phillips (1973), Phillips (1974), Robinson (1977) and Sims (1971).

We provide new results in the context of MIDAS regressions, showing that under certain conditions, the aggregation bias disappears when Y_t remains sampled at a fixed rate and only

[3]The analysis in this paper is confined to equally spaced data, albeit with different sampling frequency. The idea of MIDAS regressions can be extended to unequally spaced data, see Ghysels, Santa-Clara, Sinko, and Valkanov (2003) for further details.

[4]MIDAS regressions are obviously also not constrained to be either linear or univariate. Such extensions will also be discussed in the paper.

$X_t^{(m)}$ is sampled more frequently. In the traditional distributed lag literature it is assumed that both Y_t and X_t are sampled more frequently (see in particular Geweke (1978)). Data collection limitations prevent us often from sampling all series more frequently, hence the interest in MIDAS regressions and the interest in knowing what happens to discretization biases when only independent variables can be sampled more frequently. We show that the discretization bias in distributed lag models and in MIDAS regression both converge to zero as $m \to 0$ both in a local and global sense. This result is of significance as for instance regressions involving macroeconomic variables and financial series are usually confined to monthly, quarterly or annual regressions due to the availability of macro series. The results show that one can use the finer sampling of financial series to alleviate the discretization bias despite the unavailability of high frequency data for Y_t.

We also study the asymptotic distribution of estimators in the context of MIDAS regressions and compare them with distributed lag models. MIDAS regression parameter estimation using feasible GLS is compared with the feasible GLS in distributed lag regressions using the same regressors. We show that MIDAS regressions may be at a disadvantage in terms of asymptotic efficiency as the lack of sampling Y_t more frequently generally results in efficiency losses. Intuitively, feasible GLS can be computed using lagged dependent variables. In the conventional distributed lag model it is assumed that both $Y_t^{(m)}$ and $X_t^{(m)}$ are available, whereas in a MIDAS regression only Y_t and $X_t^{(m)}$ are. This puts constraints on the feasibility of GLS estimation.

In contrast we show that MIDAS regressions projecting Y_t onto $X_t^{(m)}$ will always improve efficiency when compared to distributed lag models involving projections of Y_t onto X_t (that is $X_t^{(1)}$). Hence, the common practice of aggregating all the data to the common least frequently sampled process will always be less efficient than a MIDAS regression that exploits the availability of $X_t^{(m)}$. This result is again of significance as for instance regressions involving macroeconomic variables and financial series are usually confined to monthly, quarterly or annual regressions due to the availability of macro series. We show that there are efficiency gains from using the finer sampled series. Finally, we also discuss various extensions of MIDAS to nonlinear and multivariate settings.

The paper is organized as follows. In section 2 we motivate the study of MIDAS regressions and discuss some of the outstanding issues. In section 3 we compare MIDAS and distributed lag models, emphasizing similarities and differences. First we revisit aggregation bias and aliasing. We are concerned with consistency, or absence of discretization bias as we sample

regressors at ever increasing frequency and show that both distributed lag and MIDAS regressions share the same properties, namely, the discretization bias is eventually eliminated. The analysis only deals with OLS estimators and does not address any efficient estimation methods. Next we study the asymptotics of MIDAS regression parameter estimation using feasible GLS and comparing it with the feasible GLS in distributed lag regressions. We show that under some special circumstances, there are no losses of efficiency when MIDAS regressions are compared with distributed lag models. We also examine situations where there are clear gains in efficiency, compared to conventional distributed lag models. The section concludes with a discussion of some similarities regarding model selection and parameterization. Section 4 deals exclusively with MIDAS models and discusses various aspects of large sample theory. The paper concludes with section 5 which lays out possible extensions and future work.

2 Why MIDAS Regressions?

MIDAS regressions are essentially tightly parameterized, reduced form regressions that involve processes sampled at different frequencies. In this section we explain why we are interested in such a setup. Consider the simple linear MIDAS regression:

$$Y_t = \beta_0 + \beta_1 B(L^{1/m}) X_{t-1}^{(m)} + \varepsilon_t^{(m)} \tag{2.1}$$

where $B(L^{1/m}) = \sum_{j=0}^{j^{max}} B(j)L^{j/m}$ is a polynomial of length j^{max} in the $L^{1/m}$ operator, and $L^{j/m}x_t{=}x_{t-j/m}$. In other words, the $L^{j/m}$ operator produces the value of x_t lagged by j/m periods.[5] Specification (2.1) is kept as simple as possible for clarity of exposition. Later sections will present more general MIDAS regressions, allowing for other regressors with different sampling frequencies, as well as multivariate and possibly non-linear relations. The order of the polynomial $B(L^{1/m})$ is assumed here (for simplicity) to be finite. However, even if the number of parameters b_k's in the polynomial $B(L^{1/m})$ is finite, it might be quite large. To capture daily fluctuations in the process over the last, say, 6 months, we would need to estimate 6×22, or 132 b_k parameters (assuming 22 trading days a month). To account for daily data over the last year, we would need approximately 264 parameters. It becomes rapidly clear that one must impose some structure upon the b_k's in order to get sensible

[5]To identify the parameter β_1 we assume that the weights of the polynomial $B(L^{1/m})$ sum to one.

results.

In empirical work, a direct treatment of mixed data samples is typically circumvented by first aggregating the highest frequency data in order to reduce all data to the same frequency. Then, in a second step, a standard regression model is estimated with pre-filtered data. This amounts to imposing some a priori restrictions on the parameters of the $B(L^{1/m})$ polynomial and by the same token not fully exploiting all the available information in the high frequency data set.

The above remarks lead us to an obvious conclusion: we face a trade-off. The mixed data sampling regression exploits a much larger information set and is more flexible. The cost is parameter proliferation, as a suitable polynomial $B(L^{1/m})$ might involve many lags of the $X_{t-j/m}^{(m)}$ data and thus many parameters to estimate. Ideally, we want to preserve most of the information in the MIDAS regression, while decreasing the number of parameters to estimate. While there are several ways of reducing the parameter space, we use an approach that is both simple to use and also is likely to suit many applications. Our approach has its roots in an old literature on distributed lag models with new twists, which is a subject that we turn to in the next section.

One may still wonder whether it is necessary to use polynomials like the ones presented in (2.1). In some cases one can indeed formulate a time series model for the data sampled at frequency $1/m$ and compute the implied MIDAS regression, an exercise called reverse engineering and discussed in Ghysels, Santa-Clara, Sinko, and Valkanov (2003). They show that in some cases temporal aggregation is possible and yields a simple mapping from the high frequency data model to the implied MIDAS regression. One often cited example is the use of the Kalman filter which is used to interpolate "missing data" (see e.g. Harvey and Pierse (1984) and subsequent work). The Kalman filter applies the linear Gaussian systems. In general settings, however, aggregation and interpolation is not so straightforward.[6] MIDAS

[6]The examples showing the complexity of reverse engineering appearing in Ghysels, Santa-Clara, Sinko, and Valkanov (2003) are drawn from the volatility literature (hence involving nonlinear models), in part because the idea of MIDAS regressions has been applied already in a number of settings involving volatility dynamics. Ghysels, Santa-Clara, and Valkanov (2002) show that MIDAS regressions for volatility provide a versatile and powerful tool to study the risk-return trade-off and improve upon existing models of volatility. See also Wang (2003) for a Bayesian model comparison which includes MIDAS regression specifications for the risk-return trade-off. Brown and Ferreira (2003), Ghysels, Santa-Clara, and Valkanov (2003), among others, use various MIDAS regressions to predict future volatility. The relationship between various recently introduced high-frequency data estimators, see e.g. Andersen, Bollerslev, Diebold, and Labys (2003) and Barndorff-Nielsen and Shephard (2003), and MIDAS regression is also discussed in Ghysels, Santa-Clara, Sinko, and Valkanov (2003).

regressions amount to a reduced form approach which avoid the impractical alternative involving the complexity of reverse engineering.

3 MIDAS and Distributed Lag Models: A Comparison

In this section we compare MIDAS and distributed lag models. We focus mostly on linear models and emphasize the differences and similarities between the two approaches. We begin with a setup where we leave unspecified the parameterization of the polynomials both for MIDAS and the distributed lag specification. In a first subsection, we revisit aliasing and discretization biases. The second subsection is devoted to asymptotic efficiency comparisons. A final subsection deals with similarities between MIDAS and distributed lag regression models.

3.1 Aggregation Bias and Aliasing Revisited

When data of different sampling frequencies are mixed, one invariably deals with temporal aggregation. To study aggregation issues it is convenient to assume that the underlying stochastic processes evolve in continuous time and data are collected at discrete points in time. Such a setting has the appeal of imposing a priori a structure on discretely observed data that is independent of the sampling interval. This is most convenient not only to study temporal aggregation but also to introduce a formal discussion of MIDAS models. Throughout the paper we shall use the convention that processes in discrete time sampled at equidistant points separated by a step size of $1/m$, are denoted by $Y_t^{(m)}$ whereas continuous time processes are denoted by $y(t)$. With this convention, observations of processes in discrete time with sampling frequency $1/m$ are:

$$Y_{k/m}^{(m)} = y(k/m) \quad \text{and} \quad X_{k/m}^{(m)} = x(k/m) \quad k \in \ldots, -1, 0, 1, \ldots \tag{3.1}$$

where $y(t)$ and $x(t) = (x_1(t), \ldots, x_N(t))'$, or more formally $y(t, \omega)$ and $x(t, \omega) = (x_1(t, \omega), \ldots, x_N(t, \omega))'$, are realizations of covariance stationary processes in continuous time

governed by a probability space (Ω, A, P).[7] The above case covers a point sampling scheme. Alternatively,

$$Y^{(m)}_{k/m} = \int_{(k-1-a)/m}^{k/m} y(\tau)d\tau \quad \text{and} \quad X^{(m)}_{k/m} = \int_{(k-1-a)/m}^{k/m} x(\tau)d\tau \tag{3.2}$$

where typically $a = 0$, though it can be positive if some type of filtering occurs (to be discussed later). The case of $m = 1$ corresponds to the discrete time representation usually studied. The superscript will often be dropped in such a case, namely Y_k refers to $Y^{(1)}_k$.

To discuss many issues ranging from parameterization and approximations to discretization biases let us start with the continuous time setting:

$$\begin{aligned} y(t) &= b * x(t) + u(t) \\ &= \int_{-\infty}^{\infty} x(t-s)b(s)ds + u(t) \end{aligned} \tag{3.3}$$

where the symbol $*$ denotes the convolution operator. The errors in equation (3.3) are not necessarily i.i.d. Identification of b in equation (3.3) rests on the assumption that the x process is, up to second moments, truly exogenous, i.e. $E[x(t)u(s)] = 0$, $\forall$ s and t.

Sims (1971) and Geweke (1978) examine equations like (3.3) and study the relationship between inference drawn from discrete time models and the parameters of the continuous time convolution.[8]

A discrete time *distributed lag* model corresponding to (3.3) would be as follows:

$$Y^{(m)}_{t/m} = \frac{1}{m} \sum_{s=-\infty}^{\infty} B^{(m)}(\frac{s}{m}) X^{(m)}_{(t-s)/m} + U^{(m)}_{t/m} \tag{3.4}$$

where both y and x are sampled at frequency $1/m$.[9] The topic of discretization bias in distributed lag models, i.e. the difference between an estimator $B^{(m)}$ and b for any given m, has been extensively studied, see for instance Sims (1971), Geweke (1978), Hansen and

[7] Further technical assumption will need to be imposed on the stochastic processes, but for the moment we shall proceed without the technical details.

[8] Equation (3.3) subsumes special cases like one-sided projections or solutions to stochastic differential equations, see e.g. Geweke (1978)

[9] The normalization of equation (3.4) by a factor $1/m$ is, as Geweke (1978) notes, necessary as the number of parameters in any set $\left[B^{(m)}(s/m) : s \in [t_1, t_2]\right]$ increases approximately in proportion with n and each individual coefficient in (3.4) will approach zero with increasing m.

Sargent (1983), Hansen and Sargent (1991b), Phillips (1972), Phillips (1973) and Phillips (1974), among others.

MIDAS regressions involve processes with various sampling frequencies. More specifically, we study projections of Y sampled with $m = 1$ and $X^{(m)}$ sampled with $m > 1$. MIDAS regression models are therefore:

$$Y_t = \frac{1}{m} \sum_{s=-\infty}^{\infty} \bar{B}^{(m)}(\frac{s}{m}) X^{(m)}_{(t-s)/m} + U_t \tag{3.5}$$

Note the differences between the two equations (3.4) and (3.5). The former has a projection of $Y^{(m)}_{t/m}$ onto the x process sampled discretely at frequency $1/m$ whereas the latter has a projection of $Y^{(1)}_t$ onto the same information set. What we are concerned with here is the comparison between $\bar{B}^{(m)}$ and $B^{(m)}$ ignoring in both cases the kind of parameterizations one wishes to impose to implement the regressions in practice.

We will consider a single regressor (as in Sims (1971)) while focusing on the limiting behavior of the discretely sampled model, as in Geweke (1978). The case of multivariate regression is, except for one key issue, a straightforward extension omitted here to avoid the cost of cumbersome notation. The key issue regarding multiple regressors omitted here, pertains to temporal aggregation which can lead to contamination across the various regressors. Geweke (1978) discusses this in great detail for the case where the dependent variable and the regressors are sampled at the same frequency. Geweke (1975) (Chapter 8) discusses cases where the regressors are sampled at different frequencies. This is a setting, as noted earlier, that features similarities with MIDAS regressions. Geweke (1975) studies in particular the cross-regressor contamination when sampling frequencies differ. Here, we focus on a single regressor and on the case where the dependent variable is not sampled more frequently.[10]

In this section we revisit the convergence of parameter estimators $B^{(m)}$ to b in equation (3.3) for m increasing in the context of a MIDAS regression model (3.5). It is important to note that we only deal with OLS estimators, and therefore are not interested at this stage with efficiency issues. The latter will be the topic of the next section. Hence, we examine OLS estimators $B^{(m)}$ in distributed lag models, similar to Sims (1971) and Geweke (1978), and OLS estimators $\bar{B}^{(m)}$ in MIDAS regressions.

[10]The multivariate setting raises other issues as well, such as testing for Granger causality. Those are discussed at length in Ghysels, Santa-Clara, and Valkanov (2003).

To do so, let us recall first what happens when a distributed lag model is considered. Following Sims (1972) the least squares estimator of $B^{(m)}$ in (3.4) minimizes the following criterion:

$$\int_{-\pi m}^{\pi m} |\tilde{B}^{(m)}(\omega) - \tilde{b}(\omega)|^2 F_m[S_x](\omega) \tag{3.6}$$

where S_x is the spectral density of the continuously sampled process $x(t)$ and the spectral density of the discretely sampled process $x_{(t-s)/m}$, denoted $S_x^{(m)} \equiv F_m[S_x]$, is expressed in terms of the folding operator (see e.g. Fishman (1969), p. 38) $F_m[g](\omega) = \sum_{k=-\infty}^{\infty} g(\omega + 2m\pi k)$. Finally, $\tilde{B}^{(m)}$ and $\tilde{b}$ are the Fourier transforms of $B^{(m)}$ and b respectively. Moreover, the discretely sampled distributed lag regression yields the OLS estimator:

$$\tilde{B}^{(m)} = F_m[S_x\tilde{b}]/F_m[S_x] = F_m[S_{yx}]/F_m[S_x] \tag{3.7}$$

where S_{yx} is the cross-spectrum of continuously sampled $y(t)$ and $x(t)$. Both equations (3.6) and (3.7) suggest that MIDAS regressions may have properties regarding discretization bias reduction similar to those of distributed lag models. Equation (3.6) tells us that the least squares estimator minimizes a least squares distance between the Fourier transform of the continuous sampling convolution polynomial and its discrete sampling fit weighted by $F_m[S_x]$. With MIDAS regressions we *do* have $F_m[S_x]$ available.

Equation (3.7) also suggests that MIDAS regressions may resemble distributed lag models in terms of discretization bias, yet it also brings us to a first technical issue that needs to be discussed. So far we did not make a distinction between discrete data driven by a point-sampling scheme, as in (3.1), or a flow aggregation as in (3.2). Usually in distributed lag models the distinction is not important. A well known result often exploited in the literature on seasonality tells us that as long as y_t and x_t are filtered with the same filter, there should be no concern regarding bias.[11] In the context of MIDAS regressions, point sampling is the most straightforward case to discuss and will therefore be treated first. When flow variables are considered one would indeed expect to see $y_t = \int_{(t-1-a)}^{t} y(\tau)d\tau$ and $x_{k/m} = \int_{(k-1-a)/m}^{k/m} x(\tau)d\tau$, which amounts to unbalanced filtering on both sides of the MIDAS regression and therefore a potential source of bias. It is for this reason that we proceed first with the point sampling case.

To proceed with the intuition why equation (3.7) also suggests that MIDAS regressions may

[11]The same filter means that a is the same in (3.2). See Sims (1974) and Wallis (1974) for the original work on the topic and Ghysels and Osborn (2001) for the most recent literature.

resemble distributed lag models in terms of discretization bias, it is important to note that what matters, besides $F_m[S_x]$, is the covariance $F_m[S_{yx}]$. In a MIDAS regression, assuming stationarity and point sampling of y and x it is clear that ultimately we recover the covariance between y_t and *any* lag of x_t. In this regards we are in a situation similar to a distributed lag model where the sampling frequency increases. There is another way to explain why distributed lag models and MIDAS regressions share similar properties with regards to discretization bias. In the previous section we noted that MIDAS regressions appear like skip-sampled distributed lag models (again thinking of the point sampling case). The skip sampling causes autocorrelated residuals, yet this does not preclude OLS to be consistent and feature the same bias properties as distributed lag models. To elaborate further on this topic we discuss the technical issues in the remainder of this section.

There is both a local and a global dimension to the bias issue, the former being point-wise $\lim_{m\to\infty} B^m(s) = b(s)$, whereas the latter is concerned how $B^m(s)$ approximates $b(s)$ as a function in the limit. It is convenient to use spectral analysis, as mean square convergence in the frequency domain is L^2 convergence in the time domain, whereas L^1 convergence in the spectral domain corresponds to point-wise convergence in the time domain. Regarding global convergence properties, Geweke (1978) (Theorem 3) shows that $\lim_{m\to\infty} \sum_{s=-\infty}^{\infty} [B^m(s/m) - b(s/m)]'[B^m(s/m) - b(s/m)] = 0$. To state the result in general terms for MIDAS regressions we consider multivariate regressions as in the original formulation of Geweke (1978). The following result can be stated as an extension of Geweke (1978) (Theorems 3 and 4):

Theorem 3.1 *Let Assumptions A.1 through A.4 appearing in Appendix A hold. Moreover, consider the MIDAS regression (3.5) with data discretely point-sampled as in (3.1), then:*

$$\lim_{m\to\infty} \sum_{s=-\infty}^{\infty} [B^m(s/m) - b(s/m)]'[B^m(s/m) - b(s/m)] = 0 \tag{3.8}$$

and for each point t there exists a sequence of intervals $S_m = (t - t_m, t + t_m)$ such that

$$\lim_{m\to\infty} (2t_m)^{-1} \sum_{s/m\in T^m} B_i^m(s/m) = \frac{1}{2} \lim_{\varepsilon\to 0} [b_i(t-\varepsilon) + b_i(t-\varepsilon)] \quad i = 1, \ldots, N \tag{3.9}$$

The proof of Theorem 3.1 appears in Appendix B. Regressions involving macroeconomic variables and financial series are usually confined to monthly, quarterly or annual regressions

due to the availability of macro series. The results appearing in this section show that one can use the finer sampling of financial series to alleviate the discretization bias.

So far we only dealt with point sampled processes and noted that flow variables are likely to be more cumbersome in the case of MIDAS regressions since mixed sampling frequencies lead to different flow aggregations. Recall that the cause of the problem is the unbalanced filtering $y_t = \int_{(t-1-a)}^{t} y(\tau)d\tau$ and $x_{k/m} = \int_{(k-1-a)/m}^{k/m} x(\tau)d\tau$. There is, however, a fairly simple - although somewhat unorthodox - solution to the bias induced by unbalanced filtering. It suffices to project y_t onto $x_{k/m} = \int_{(k/m-1-a)}^{k/m} x(\tau)d\tau$, which amounts to a balanced filtering on both sides of the MIDAS regression. This scheme yields a MIDAS regression where for example quarterly GNP growth is projected on monthly sampled 3-month inflation growth rates. Likewise, in the case of volatility applications this scheme would amount to projecting daily increments in quadratic variation onto five-minute sampled daily increments in quadratic variation (assuming a 24-hour market cycle).[12]

To conclude this section we would like to draw attention to the dimensionality of aliasing, as discussed in Hansen and Sargent (1983). In the case of rational polynomial lags Hansen and Sargent (1983) (Theorem 1) show that in general there will only be finite number of observationally equivalent models due to aliasing (though in general the class of observationally equivalent models given equispaced discrete time series observations is uncountable). Their result readily applies to MIDAS regressions as well.

3.2 Asymptotic Efficiency

The asymptotic analysis in the previous section was one of continuous records and the emphasis was consistency, or absence of discretization bias as we sampled regressor at ever increasing frequency. We showed that both distributed lag and MIDAS regressions feature the desirable property of approximating b both locally and a globally. Moreover, the analysis in the previous section only dealt with OLS estimators and did not address any efficient estimation methods. In this section we turn our attention to efficient estimation. To do so, we turn our attention to the conventional asymptotic analysis where the *span* of the data set T expands asymptotically with a fixed sampling frequency m. Distributed lag models will have sample sizes mT whereas the corresponding sample sizes for MIDAS regressions

[12]Such a scheme has been considered in the context of volatility estimation by Andreou and Ghysels (2002) as a rolling sample estimator of increments of quadratic variation.

will be T. Obviously, with $m = 1$ both are equivalent and MIDAS regressions turn into distributed lag models. Consequently, distributed lag models involve more 'data' as the number of observations is mT, yet as far as *information set* is concerned, both distributed lag and MIDAS regressions are on equal footing since they both involve the same regressors. What we are interested in is what happens as $T \to \infty$ so that both samples are large and involve the same regressors.

We begin our analysis with linear models, which build directly on the discussions appearing in the previous section. Linear models are covered in a first subsection. Next, we move to partial linear models which feature nonlinearities separable from a linear projection and therefore share many properties with linear models. A third and final section deals with general nonlinear models.

3.2.1 Linear Regression Models

As in the previous section, it is not surprising that we will rely on spectral estimation and in particular examine estimators due to Hannan (1963a) and Hannan (1963b) that are asymptotically normal and efficient. The frequency domain GLS achieves asymptotically the Gauss-Markov efficiency bound under general smoothness conditions on the residual spectral density.

Consider again the discrete time *distributed lag* model like (3.4) where both y and x are sampled at a *fixed* frequency $1/m$. Hence, we consider equation

$$Y_{t/m}^{(m)} = \frac{1}{m} \sum_{s=-\infty}^{\infty} b^{(m)}(\frac{s}{m}) X_{(t-s)/m}^{(m)} + u_{t/m}^{(m)} \tag{3.10}$$

where $b^{(m)}$ is the pseudo-true value associated with the fixed m.[13] We try to obtain an *efficient* estimator which we will denote B_H^m given a data set of size mT for both $Y^{(m)}$ and $X^{(m)}$.

[13]Note the two differences between equations (3.4) and (3.10). The latter uses pseudo-true parameters $b^{(m)}$ and residuals $u_{t/m}^{(m)}$, whereas the former was expressed in terms of OLS estimator $B^{(m)}$ and estimated residuals $U_{t/m}^{(m)}$.

Before discussing the asymptotic distribution for B_H^m we introduce the MIDAS regression:

$$Y_t = \frac{1}{m}\sum_{s=-\infty}^{\infty} \bar{b}^{(m)}(\frac{s}{m})X_{(t-s)/m}^{(m)} + u_t \tag{3.11}$$

where $\bar{b}^{(m)}$ is again the pseudo-true value associated with the fixed m in analogy with equation (3.10). The efficient estimator for the above MIDAS regression, which we will denote B_M^m given a data set of size T for Y and $X^{(m)}$ has the following properties. The efficient estimator for the above MIDAS regression, which we will denote B_M^m given a data set of size T for Y and $X^{(m)}$ has the following properties, in comparison with the distributed lag model estimator B_H^m :

Theorem 3.2 *Let Assumptions A.1 through A.6 appearing in Appendix A hold. Then, the Hannan feasible estimator is defined as:*

$$\hat{B}_H^{(m)} = [\sum_{j=-k_m+1}^{k_m} \hat{S}_X^{(m)}(\omega_j)\hat{S}_U^{(m)}(\omega_j)^{-1}]^{-1}[\sum_{j=-k+1}^{k} \hat{S}_{XY}^{(m)}(\omega_j)\hat{S}_U^{(m)}(\omega_j)^{-1}] \tag{3.12}$$

for $\omega_j = m\pi j/k_m$ and where the spectral density estimators and bandwidth are defined in (A.1) appearing in Appendix A. Likewise, the Hannan feasible estimator for a MIDAS regression is:

$$\hat{B}_M^{(m)} = [\sum_{j=-k_m+1}^{k_m} \hat{S}_X^{(m)}(\omega_j)\hat{S}_U^{(1)}(\omega_j)^{-1}]^{-1}[\sum_{j=-k+1}^{k} \hat{S}_{XY}^{(m)}(\omega_j)\hat{S}_U^{(1)}(\omega_j)^{-1}] \tag{3.13}$$

The estimator (3.12) has the following asymptotic distribution:

$$\sqrt{mT}(\hat{B}_H^{(m)} - b^{(m)}) \to N(0, 2\pi\{\int_{-m\pi}^{m\pi} F_m[S_x(\omega)](F_m[S_u(\omega)])^{-1}d\omega\}^{-1}) \tag{3.14}$$

whereas estimator (3.13) has the following asymptotic distribution:

$$\sqrt{T}(\hat{B}_M^{(m)} - \bar{b}^{(m)}) \to N(0, 2\pi\{\int_{-\pi}^{\pi} F_m[S_x(\omega)](F_1[S_u(\omega)])^{-1}d\omega\}^{-1}) \tag{3.15}$$

Provided, $\bar{b}^{(m)}$ and $b^{(m)}$ are equal, the two estimators are asymptotically equivalent if $F_m[S_u]$ is constant, i.e. $U^{(m)}$ is white noise.

The proof of the above theorem appears in Appendix C. Note that the pseudo-true values $\bar{b}^{(m)}$ and $b^{(m)}$ might differ, although the results of the previous section warrant to assume that such a difference would be negligible for sufficiently large m. Therefore, in the remainder of our analysis we will ignore any differences that might exist between $\bar{b}^{(m)}$ and $b^{(m)}$.

Let us first further elaborate on why the asymptotic efficiency of distributed lag and MIDAS regressions differ. To do this it will be helpful to consider a slight variation of equation (3.3). Often the equation is obtained from a so called rational distributed lag:

$$b_2 * y(t) = b_1 * x(t) + v(t) \tag{3.16}$$

where identification of b_1 and b_2 is achieved by assuming that v is serially uncorrelated as well as uncorrelated with x.

Equations (3.16) and (3.3) are related via the relationship $b \equiv b_2^- * b_1$ where b_2^- is the inverse under convolution. Consequently, the serial dependence of the residuals in (3.3) is determined by $v(t) = b_2^- * u(t)$. A discrete time distributed lag model corresponding to (3.16) would be as follows:

$$Y_{t/m}^{(m)} = \frac{1}{m} \sum_{s=-\infty}^{\infty} b_1^{(m)}(\frac{s}{m})(b_2^{(m)}(\frac{s}{m}))^- X_{(t-s)/m}^{(m)} + u_{t/m}^{(m)}$$

A simple strategy that leads to efficient estimation is to prefilter the equation by b_2 :

$$Y_{t/m}^{(m)} = \sum_{s=-\infty}^{\infty} (\tilde{b}_2^{(m)}(\frac{s}{m}))Y_{(t-s-1)/m}^{(m)} + \sum_{s=-\infty}^{\infty} b_1^{(m)}(\frac{s}{m})X_{(t-s)/m}^{(m)} + v_{t/m}^{(m)}$$

where the availability of lagged $Y_{t/m}^{(m)}$ allows us to apply the polynomial b_2. In a MIDAS regression this strategy is infeasible due to the lack of high frequency $Y_{t/m}^{(m)}$. Consequently, the errors remain correlated and the estimator has to settle with an autocorrelation structure that cannot be further unravelled. The advantage of distributed lag models is the availability of the additional information about $Y^{(m)}$.

The result in Theorem 3.2 tells us that uncorrelated errors in the distributed lag equation are a situation where the advantage of distributed lag models is not of any consequence as there is no need to prefilter. This observation is valid for models that are not determined by rational polynomials as well, the case of rational polynomials is one where the results can be presented in a transparent way. It is important to note, however, that theorem 3.2 does not state that white noise is both necessary and sufficient. Indeed, there are cases where the two

estimators are asymptotically equally efficient despite the fact that $F_m[S_u]$ is not constant, i.e. $U^{(m)}$ is autocorrelated. A simple case would be where $U^{(m)}$ is an MA(q) process with $q < m$. In such situations, there is correlation in $U^{(m)}$ but $U^{(1)}$ is uncorrelated as the original process has memory shorter than the temporal aggregation. The Hannan efficient estimator of the distributed lag model picks up the autocorrelation up to lag q, whereas the MIDAS regression is asymptotically efficient without such a correction.

There is another important comparison between MIDAS and distributed lag models that needs to be considered. As noted before, a direct treatment of mixed data samples is typically circumvented by first aggregating the highest frequency data in order to reduce all data to the same frequency. Then, in a second step, a standard regression model is estimated with pre-filtered data. In such a case one should consider a distributed lag model where both y and x are sampled at a *fixed* low frequency $m = 1$. Hence, we consider equation

$$Y_t^{(1)} = \sum_{s=-\infty}^{\infty} b^{(1)}(s) X_{t-s}^{(1)} + u_t^{(1)} \tag{3.17}$$

The above equation is compared with (3.11) where the same Y_t is projected onto the high frequency set of regressors instead of aggregating them. The efficient estimator for the MIDAS regression, which we will again be denoted B_M^m given a data set of size T for Y and $X^{(m)}$ has the following properties, in comparison with the distributed lag model estimator $B_H^{(1)}$:

Theorem 3.3 *Let Assumptions A.1 through A.6 appearing in Appendix A hold. Then, the Hannan feasible estimator $\hat{B}_H^{(1)}$ as defined in (3.12) with $m = 1$ has the asymptotic distribution:*

$$\sqrt{T}(\hat{B}_H^{(1)} - b^{(1)}) \to N(0, 2\pi\{\int_{-\pi}^{\pi} F_1[S_x(\omega)](F_1[S_u(\omega)])^{-1} d\omega\}^{-1}) \tag{3.18}$$

whereas the MIDAS regression estimator (3.13) has the asymptotic distribution appearing in (3.15). Provided, $\bar{b}^{(m)}$ and $b^{(1)}$ are equal, the MIDAS regression estimators is asymptotically more efficient.

The proof of the above theorem follows from the fact that for each frequency $\omega \in [-\pi, \pi]$, $F_1[S_x(\omega)] \geq F_m[S_x(\omega)]$. The latter property holds by the definition of the folding operator, or equivalently, as a consequence of aliasing due to aggregation of the regressors. The

implications of Theorem 3.3 are important as they tell us that MIDAS regressions are more efficient than the common practice of first aggregating the highest frequency data in order to reduce all data to the same frequency.

To conclude we should first of all note that Hannan's estimation procedure requires the choice of a bandwidth k_m, and an unsuitable bandwidth selection can produce poor estimates. Robinson (1991) discusses frequency domain inference with data-based bandwidth selection and proposed a commonly used spectral estimator based on a weighted average of periodogram estimates of the fundamental frequencies, or:

$$\hat{B}_R^{(m)} = [\sum_{j=-mT/2+1}^{mT/2} I_X^{(m)}(\omega_j)\hat{S}_U^{(m)}(\omega_j)^{-1}]^{-1}[\sum_{j=-mT/2+1}^{mT/2} I_{XY}^{(m)}(\omega_j)\hat{S}_U^{(m)}(\omega_j)^{-1}] \tag{3.19}$$

The above estimator $\hat{B}_R^{(m)}$ is first order equivalent to the original estimator proposed by Hannan. It is not difficult to show that the results in this section extend to such alternative estimators when MIDAS and distributed lag regressions are compared in terms of asymptotic efficiency. One outstanding issue, beyond the scope of the present paper is how higher-order approximations for the coefficient estimates in MIDAS and distributed lag models compare. Xiao and Phillips (1998) discuss such expansions for $\hat{H}_R^{(m)}$. We leave such analysis for future research.

Finally, it should also be noted that simultaneous equations linear MIDAS regressions can also be studied and compared with systems of linear distributed lag regressions. Indeed, the analysis in this section, using the Hannan efficient estimation procedure, has multivariate extensions. In particular, Hannan (1968) studies the circumstances under which least squares are asymptotically efficient for the estimation of in systems of linear regressions and provides a theorem which can be used to extend the result in Theorem 3.2 to multivariate settings.

3.2.2 Partial Linear Models

The analysis in this section is inspired by Phillips, Guo, and Xiao (2002) who consider:

$$Y_{t/m}^{(m)} = \frac{1}{m}\sum_{s=-\infty}^{\infty} b^{(m)}(\frac{s}{m})X_{(t-s)/m}^{(m)} + g(Z_{t/m}^{(m)}) + u_{t/m}^{(m)} \tag{3.20}$$

where the above equation is an adaptation of (3.10) to include a nonlinear functional g.[14] Hence, in this model the response is assumed to be linearly related to $X_{t/m}^{(m)}$ and nonlinearly to $Z_{t/m}^{(m)}$ (without lags). Partial linear models have been studied extensively and Phillips, Guo, and Xiao (2002) provide an elaborate list of papers on the subject. Following early work by Robinson (1988), a Nadaraya-Watson kernel estimator is used to eliminate the unknown nonlinear function in a first step. Robinson (1988) assumed i.i.d. errors and showed that a second stage least squares estimator for the linear regression part is $\sqrt{mT}$ consistent and asymptotically normal. Phillips, Guo, and Xiao (2002) extends this to general autocorrelated residuals and use a spectral density approach like in the previous section. Consider the following MIDAS partial linear regression:

$$Y_t = \frac{1}{m}\sum_{s=-\infty}^{\infty} \bar{b}^{(m)}(\frac{s}{m})X_{(t-s)/m}^{(m)} + g(Z_{t/m}^{(m)}) + u_t \tag{3.21}$$

Taking expectations conditional on $Z_{t/m}^{(m)}$ in both equations (3.20) and (3.21) and subtracting the result from the original equations yields:

$$\tilde{Y}_{t/m}^{(m)} = \frac{1}{m}\sum_{s=-\infty}^{\infty} b^{(m)}(\frac{s}{m})\tilde{X}_{(t-s)/m}^{(m)} + u_{t/m}^{(m)}\tilde{Y}_t = \frac{1}{m}\sum_{s=-\infty}^{\infty} b^{(m)}(\frac{s}{m})\tilde{X}_{(t-s)/m}^{(m)} + u_t$$

where $\tilde{Y}_{t/m}^{(m)} = Y_{t/m}^{(m)}$ - $E[Y_{t/m}^{(m)}|Z_{t/m}^{(m)}]$, $\tilde{Y}_t = Y_t$ - $E[Y_t|Z_t^{(m)}]$, and $\tilde{X}_{t/m}^{(m)} = X_{t/m}^{(m)}$ - $E[X_{t/m}^{(m)}|Z_{t/m}^{(m)}]$. Note that $\tilde{Y}_t$ is still conditional on the same $Z^{(m)}$ process as $\tilde{Y}_{t/m}^{(m)}$. If the conditional expectations were known, the above regression would simply be respectively a linear distributed lag and MIDAS regression. In partial linear models the quantities $\tilde{Y}_{t/m}^{(m)}$, $\tilde{Y}_t$ and $\tilde{X}_{t/m}^{(m)}$ involve nonparametric estimation using a standard Nadaraya-Watson kernel estimator. The analysis of Robinson (1988) and Phillips, Guo, and Xiao (2002) allows us to extend theorem 3.2 to partial linear MIDAS models.[15]

[14] To be precise Phillips, Guo, and Xiao (2002) consider a regression such as (3.20) with a general regressor which we have specialized to the distributed lag setting.

[15] It should be noted, however, that the technical assumptions appearing in Appendix A require some strengthening, see Phillips, Guo, and Xiao (2002) for details.

3.3 Some Similarities

The most striking similarity between MIDAS regressions and distributed lag models is the fact that lag polynomials need to be tightly parameterized. In this respect there are similar issues that emerge. Various parameterizations have been suggested in the distributed lag literature, see e.g. Judge, Griffith, Hill, Lutkepohl, and Lee (1985) for further discussion.[16] This common theme between distributed lag and MIDAS regressions generates similarities with regards to estimation. Take for example a "rational" polynomial lag structure, as appearing in equation (3.16). Often such a rational polynomial is thought of as an approximation for the function $b(s)$ in (3.3). Therefore, model selection issues and asymptotic misspecification errors are relevant for both MIDAS and distributed lag regressions and there is no new theory as far as MIDAS is concerned. Spectral estimation typically amounts to fixing the model size deterministically as a function of the sample size (see Sims (1974) for further discussion). In a different approach, due to Akaike (1973) and many subsequent refinements such as Schwarz (1978), among many others, a model fitting information criterion function is used. We do not further explore this area here, except for noting that there is a large literature already on the subject that can be applied in the context of MIDAS regressions.

4 General MIDAS Models

It will be convenient to start from a conventional asymptotic analysis. Let us consider a general multivariate MIDAS regression setting, namely:

$$Y_{t+1} = \mathcal{B}_0 + f(\sum_{i=1}^{K}\sum_{j=1}^{L} \mathcal{B}_{ij}(L^{1/m_i})g(X_t^{(m_i)}, \beta), \beta) + \varepsilon_{t+1} \tag{4.1}$$

and we collect all the parameters controlling the polynomials into the parameter vector b. As noted in the previous section, the polynomials $\mathcal{B}_{ij}(L^{1/m_i})$ can be two-sided and the functions f and/or g can involve unknown parameters β. The parameters that determined the polynomial lag weights are collected into a parameter vector b. When unconstrained

[16] Ghysels, Santa-Clara, and Valkanov (2003) introduce a distributed lag based on the beta function, which is to the best of our knowledge novel to the literature and has proven to be very useful. The lag structure can take many shapes and is determined only by two parameters.

estimation is considered the latter is potentially infinite. In the context of MIDAS regression models the parameter vector b is a function of hyperparameters θ, therefore we will use the notation $b(\theta)$. To separate the hyperparameter vector θ controlling the polynomials from the other parameters we denote $\gamma = (\beta' \quad \theta')'$. Therefore unconstrained estimation involves the possibly infinite parameter space $(\beta' \quad b')'$, which is replaced in a MIDAS regression by $(\beta' \quad b(\theta)')'$, or $(\beta' \quad \theta')'$.[17] At first we will assume *fixed* m_i, $i = 1, \ldots, K$, and show that for such cases we can estimate MIDAS regression with the usual asymptotic tools. Hence, for all practical purposes one can do the estimation with standard software using conventional econometric methods.

The asymptotic analysis becomes slightly more involved when we let at least one m_i go to infinity, implying a continuous record conditioning set of regressors. In a first subsection we present the conventional asymptotic analysis and then in a second subsection we turn to MIDAS regressions with continuous record observations.

4.1 Fixed and Finite Sampling Frequencies

We consider the general class of extremum estimators. This class, which maximizes some objective function that depends on the data and sample size, includes maximum likelihood (MLE), nonlinear least squares (NLS) and generalized method of moments (GMM) estimators which are the three types of estimators we would like to consider. An estimator $\hat{\gamma}_T$ is an extremum estimator if there is an objective function $\hat{M}_T(\gamma)$, given a sample size T such that $\hat{\theta}_T$ maximizes $\hat{M}_T(\gamma)$ subject to $\theta \in \Gamma$. The MLE estimator corresponds to

$$\hat{M}_T(\gamma) \equiv T^{-1} \sum_{t=1}^{T} l(\varepsilon_t|\gamma) \tag{4.2}$$

where l is the log likelihood based on distributional assumptions on the error process in (4.1). As for the NLS estimator, the objective function is

$$\hat{M}_T(\gamma) \equiv -T^{-1} \sum_{t=1}^{T} \varepsilon_t(\gamma)^2 \tag{4.3}$$

[17] We keep using the term MIDAS regression, although we are now dealing potentially simultaneous equations.

where $\varepsilon_{t+1}(\gamma) \equiv [y_{t+1} - \mathcal{B}_0 - f(\sum_{i=1}^{K}\sum_{j=1}^{L} \mathcal{B}_{ij}(L^{1/m_i})g(X_t^{(m_i)}))]$. Finally for the GMM estimator the objective function

$$\hat{M}_T(\gamma) \equiv -[T^{-1}\sum_{t=1}^{T} g_t(\gamma)]'\hat{W}_T[T^{-1}\sum_{t=1}^{T} g_t(\gamma)] \tag{4.4}$$

where $g_t(\gamma) \equiv \upsilon_t \times Z_{t-1}$ where Z_{t-1} is an instrument vector.[18]

One of the standard regularity conditions for consistency is that the parameter space is compact, which in most cases is achieved by assuming a finite dimensional closed and bounded parameter space. More specifically, $\Gamma \subset \mathbf{R}^q$ and Γ is compact. MIDAS regressions therefore assume the standard environment in terms of parameter spaces. A second critical assumption to establish consistency is identification, which can be written as:

Assumption 4.1 *Given the information set $I_t \equiv X_\tau^{(m_i)}, \tau < t$, $i = 1, \ldots, K$, there exists a function $b(\theta_0)$ with dim(θ_0) finite (small) and a parameter β_0 such that*

$$E[\varepsilon_{t+1}(\beta_0', b(\theta_0))')|I_t] = 0$$

for a unique $\gamma_0 = (\beta_0', \theta_0')' \in \Gamma \subset \mathbf{R}^q$ and Γ is compact.

This assumption is critical as it ensures the correct specification of the MIDAS polynomials. When this assumption replaces the usual identification assumption we obtain the usual asymptotic results, provided all other standard regularity conditions apply. More specifically, the MLE, NLS and GMM estimators are consistent and asymptotically normal under suitable regularity conditions appearing for instance in Gallant and White (1988), among many others. Note that the asymptotics is for fixed m_i, $i = 1, \ldots, K$, and T going to infinity.

4.2 Continuously Sampled Regressors

In this section we devote our attention to cases where at least one m_i in (4.1) goes to infinity, implying a continuous record conditioning set of regressors. Hence, we ultimately estimate a functional approximation with a continuum of past observations rather than a polynomial

[18] Recall that when autoregressive augmentations appear in MIDAS regressions we know that the lagged dependent variable may not be a valid instrument, as discussed earlier.

lag of a MIDAS regression. To discuss this case, we focus on a univariate single regressor model without intercept and slope:

$$Y_{t+1} = B(L^{1/m})X_t^{(m)} + \varepsilon_{t+1}^{(m)} \tag{4.5}$$

where $B(L^{1/m}) = b_0 + b_1 L^{1/m} + b_2 L^{2/m} + \ldots + b_{j^{max}} L^{j^{max}/m}$.[19] Suppose now that we take the limit of $m \to \infty$ with $j^{max}/m \to \kappa$. Hence, we are essentially sampling a continuum of data between t and $t-\kappa$, allowing possibly κ to be infinite. With a continuum of data (4.5) becomes the following convolution equation:

$$Y_{t+1} = \beta_0 + \beta_1 \int_{j=0}^{\kappa} b_j(\theta) X_{t-j}^{(\infty)} dj + \varepsilon_{t+1}^{(\infty)} \tag{4.6}$$

The MLE and NLS estimators of a correctly specified MIDAS regression, that is one satisfying Assumption 4.1, are again standard provided we can compute the integral in (4.6) without numerical approximation error. Note that now $\varepsilon_{t+1}(\theta) \equiv y_{t+1} - \int_{j=0}^{\kappa} b_j(\theta) x_{t-j}^{(\infty)} dj$.

The GMM estimator requires more discussion because the choice of moment conditions and instruments is not so straightforward. Recall that the GMM estimator specializes to

$$\hat{M}_T(\theta) \equiv -[T^{-1} \sum_{t=1}^{T} [(y_{t+1} - \int_{j=0}^{\kappa} b_j(\theta) x_{t-j}^{(\infty)} dj) Z_{t-1}]' \hat{W}_T [T^{-1} \sum_{t=1}^{T} [(y_{t+1} - \int_{j=0}^{\kappa} b_j(\theta) x_{t-j}^{(\infty)} dj) Z_{t-1}]] \tag{4.7}$$

and in principle any $x \in I_{t,t-\kappa}^{(\infty)}$ is a valid instrument so that one can exploit all possible moment conditions that arise from the cross-product of errors and regressors in the MIDAS regression polynomial. This ultimately yields a continuum of moment conditions, with a finite parameter space. The fact that we approach a continuum of moments implies that the moment conditions in (4.7) become more correlated and in the limit their covariance matrix (and hence the inverse of the optimal GMM weighting matrix) approaches singularity. This problem has been recognized by Carrasco and Florens (2000), who propose a so called $C-GMM$ estimator in situations of a limit continuum of moment conditions.

The $C-GMM$ estimator is based on the arbitrary set of moment conditions:

$$E^{\theta_0} h_t(\tau; \theta_0) = 0 \tag{4.8}$$

[19] For simplicity we also assume that the polynomial to be one-sided.

where $h_{t+1}(\tau;\theta) \equiv [y_{t+1} - \int_{j=0}^{\kappa} b_j(\theta) x_{t-j}^{(\infty)} dj] x_{t-\tau}^{(\infty)}$, with $\tau \in \mathcal{R}^+$. We will refer to $h_t(\tau;\theta_0)$ as a moment function.[20] Let $\hat{h}_T(\tau;\theta_0) = \sum_{t=1}^{T} h_t(\tau;\theta_0)/T$ denote the sample mean of the moment functions. The most convenient way to work with such infinite set is to impose a Hilbert space structure. Carrasco and Florens introduce a space $\mathcal{L}^2(\pi)$ to which $h_t(.;\theta_0)$ belongs as a function of τ. The inner product in this space is defined as

$$\langle f, g \rangle = \int f(\tau) g(\tau) \pi(\tau) d\tau \tag{4.9}$$

where π is a probability density usually selected to be Gaussian. The norm corresponding to the inner product is $\| f \|^2 = \langle f, f \rangle$. Similar to the standard GMM setup, one can prove the central limit result for the sample mean of moment functions:

$$\sqrt{T}\hat{h}_T(\tau;\theta_0) \overset{\mathcal{L}}{\Rightarrow} \mathcal{N}(0, K) \tag{4.10}$$

Since h_T is an element of Hilbert space, $\mathcal{N}$ is understood as a Gaussian random element of the same space with variance $\langle Kf, f \rangle$, where the covariance operator K satisfies:

$$\langle Kf, g \rangle = E^{\theta_0} [\langle f, h_t(\theta_0) \rangle \langle g, h_t(\theta_0) \rangle] \tag{4.11}$$

Note that K is an integral operator that can be written as

$$Kf(\tau_1) = \int k(\tau_1, \tau_2) f(\tau_2) \pi(\tau_2) d\tau_2 \tag{4.12}$$

with $k(\tau_1, \tau_2) = E^{\theta_0}\left(h_t(\tau_1;\theta_0) h_{t(\tau_2;\theta_0)}\right)$. The function k is called the kernel of the integral operator K.

One way to implement the C-GMM estimator is to minimize the objective function:

$$\min_{\theta} \underline{\bar{v}}'(\theta) \left[I_T - C \left[\alpha_T I_T + C^2 \right]^{-1} C \right] \underline{v}(\theta) \tag{4.13}$$

where C is a $T \times T-$matrix with the eigenvalues identical to those of K_T and with (t, l)

[20] We continue here with the special case of a single regressor. Multi-regressor or multivariate extensions are straightforward extensions.

element $c_{tl}/(T-q)$, $t,l=1,...,T$, I_T is the $T\times T$ identity matrix, $\underline{v}=[v_1,...,v_T]'$ with

$$\begin{aligned} v_t(\theta) &= \left\langle \hat{h}_T(\tau;\theta), h_t\left(\tau;\hat{\theta}_T^1\right)\right\rangle, \\ c_{tl} &= \left\langle h_l\left(\tau;\hat{\theta}_T^1\right), h_t\left(\tau;\hat{\theta}_T^1\right)\right\rangle. \end{aligned}$$

where $\hat{\theta}_T^1$ is a first step estimator which consistent (as in the usual GMM setting). The above estimator, when Assumption 4.1 which guarantees that the MIDAS regression is asymptotically correctly specified, has the standard properties of GMM estimators: consistency, asymptotic normality and optimality. The following result is stated without proof, as details appear in Carrasco and Florens (2000) and Carrasco, Chernov, Ghysels, and Florens (2002):

Proposition 4.1 *Let Assumption 4.1 hold and all other regularity conditions for the C-GMM appearing in Carrasco and Florens (2000) hold as well. Moreover, let B be a bounded linear operator defined on $\mathcal{L}^2(\pi)$ or a subspace of $\mathcal{L}^2(\pi)$ and B_T a sequence of random bounded linear operators converging to B. The C-GMM estimator $\hat{\theta}_T = \underset{\theta}{argmin} \left\| B_T \hat{h}_T(\theta) \right\|$ has the following properties:*

1. *$\hat{\theta}_T$ is consistent and asymptotically normal such that*

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, V_1^{-1} \times V_2 \times V_1^{-1}\right)$$

   where $V_1 = \left\langle BE^{\theta_0}(\nabla_\theta h), BE^{\theta_0}(\nabla_\theta h)\right\rangle$ and $V_2 = \left\langle BE^{\theta_0}(\nabla_\theta h), (BKB^*)BE^{\theta_0}(\nabla_\theta h)\right\rangle$.

2. *Among all admissible weighting operators B, there is one yielding an estimator with minimal variance. It is equal to $K^{-1/2}$, where K is the covariance operator defined in (4.12).*

Carrasco, Chernov, Ghysels, and Florens (2002) extend this to the case of weakly dependent processes. If it is a weakly dependent process then, h_t is replaced by Uh_t in v_t and c_{tl}, see Carrasco, Chernov, Ghysels, and Florens (2002) for a definition of Uh_t and further details. This estimator, like the usual GMM, also involves a two-step procedure and a HAC-type estimator of the covariance operator.

It is important to stress that in the above analysis, the sample size T drives the asymptotics. This is perhaps not surprising since the left hand side of a MIDAS regression determines the

data accumulation rate in terms of the reference interval of time. In this regard, our analysis differs from recent developments such as Barndorff-Nielsen and Shephard (2003), who study a multivariate covariance and regressions framework and consider "filling in" of data $x^{(m)}$ over fixed time intervals and obtain non-Gaussian asymptotic distributions. Along these lines one could consider letting the sampling interval of Y_t and $X_t^{(m)}$ shrink at appropriate rates to yield a continuous record data sample. We leave this question open for future research.

Once a continuum of moments approach is considered one can also wonder what the most efficient choice of instruments would be. Carrasco, Chernov, Ghysels, and Florens (2002) consider so called double index moment functions where τ in (4.8) is multidimensional, that is $\tau = (\tau_1 \quad \tau_2) \in R^2$.[21] In particular, consider the set of moment conditions:

$$h_{t+1}(\tau;\theta) \equiv [y_{t+1} - \int_{j=0}^{\theta} b_j(\theta) x_{t-j}^{(\infty)} dj] Z(\tau_1, x_{t-\tau_2}^{(\infty)}) \tag{4.14}$$

where $Z(\tau_1, x_{t-\tau_2}^{(\infty)})$ is some 'optimal' instrument choice. Using results in Carrasco, Chernov, Ghysels, and Florens (2002) one can compute the asymptotic variance of $\hat{\theta}_T$, namely one can compute $\left(\left\langle E^{\theta_0}(\nabla_\theta h), E^{\theta_0}(\nabla_\theta h)\right\rangle_K\right)^{-1}$. To establish conditions under which this variance coincides with the Cramer Rao efficiency bound, consider S, the linear space spanned by $\{h(\tau, y_t; \theta_0)\}$ and $\overline{S}$ be its closure. The results in Carrasco, Chernov, Ghysels, and Florens (2002) imply that double-index C-GMM estimator based on (4.14) is efficient when the score belongs to the span of the moment conditions. Intuitively, such a choice of instrument should be clear. Since we can not construct the optimal instrument in, we can span it via a set of basis functions. The choice of functions $Z(\tau_1, x_{t-\tau_2}^{(\infty)})$ is closely related with the choice of test functions to construct consistent conditional moment test, see Bierens (1990) as well as Stinchcombe and White (1998) and references therein. In particular, using the results of Stinchcombe and White (1998), $Z(\tau_1, x_{t-\tau_2}^{(\infty)})$ could be based on any analytic functions but the polynomials. One choice would be to consider the set of base functions $Z(\tau_1, x_{t-\tau_2}^{(\infty)}) = \exp \tau_1 x_{t-\tau_2}^{(\infty)}$, with $\tau_1 \in \mathcal{R}$ and $\tau_2 \in \mathcal{R}^+$. The utilization of the continuum of moment conditions is precisely what allows one to perform this spanning. Needless to say that imposing a distributional assumption on υ_t yields an efficient MLE estimator that can be implemented straightforwardly as well. The issue of efficient estimation also needs further exploration.

[21] We continue here again with the special case of a single regressor. Multi-regressor or multivariate extensions are straightforward extensions.

5 Conclusions

We introduced MIDAS regression models which involve time series data sampled at different frequencies. MIDAS regressions are essentially tightly parameterized reduced form regressions that involve processes sampled at different frequencies. At a general level, the interest in MIDAS regressions addresses a situation often encountered in practice where the relevant information is high frequency data, whereas the quantity of interest is a low frequency process. In addition, MIDAS regressions also address situations where data collection limitations result in data that are not sampled at the same frequency. In empirical work, a direct treatment of mixed data samples is typically circumvented by first aggregating the highest frequency data in order to reduce all data to the same frequency and then in a second step estimate a standard regression model. We examined the features MIDAS regressions share with distributed lag models but also emphasized their unique novel features. Among the most important findings, we show that MIDAS regressions will always lead to more efficient estimation than the typical approach of aggregating all series to the least frequent sampling. In some cases MIDAS regressions are also as efficient as distributed lag regressions with all series available at the highest frequency. We also show that discretization biases are the same for MIDAS and distributed lag models and vanish when regressors are sampled more frequently.

While we discussed a large variety of issues, we clearly indicated some areas that remain unresolved. These areas pertain to estimation and specification errors as well as the treatment of long memory, seasonality and other common time series themes like (fractional) co-integration.

References

Akaike, H., 1973, "Information theory and an extension of the maximum likelihood principle," in *Second International Symposium on Information Theory*, ed. by B. Petrov, and F. Csaki, pp. 267–281. Akademia Kiado (Budapest).

Andersen, T., T. Bollerslev, F. X. Diebold, and P. Labys, 2003, "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

Andreou, E., and E. Ghysels, 2002, "Rolling-Sample Volatility Estimators: Some New Theoretical, Simulation and Empirical Results," *Journal of Business and Economic Statistics*, 20, 363–376.

Barndorff-Nielsen, O., and N. Shephard, 2003, "Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics," *Econometrica*, 72, 885–925.

Bergstrom, A., 1990, *Continuous Time Econometric Modelling*. Oxford University Press, Oxford.

Bierens, H., 1990, "A consistent conditional moment test of functional form," *Econometrica*, 58, 1443–1458.

Brown, D. P., and M. A. F. Ferreira, 2003, "The Information in the Idiosyncratic Volatility of Small Firms," Working paper, Univesrity of Wisconsin and ISCTE.

Carrasco, M., M. Chernov, E. Ghysels, and J. Florens, 2002, "Efficient estimation of jump diffusions and general dynamic models with a continuum of moment conditions," Discussion Paper.

Carrasco, M., and J. P. Florens, 2000, "Generalization of GMM to a continuum of moment conditions," *Econometric Theory*, 16, 797–834.

Chambers, M., 1991, "Discrete models for estimating general continuous time systems," *Econometric Theory*, 7, 531–542.

Comte, F., and E. Renault, 1996, "Noncausality in continuous time models," *Journal of Econometrics*, 73, 101–149.

Dhrymes, P., 1971, *Distributed lags: Problems of Estimation and Formulation*. Holden-Day, San Francisco.

Fishman, G., 1969, *Spectral Methods in Econometrics*. Harvard University Press, Cambridge.

Gallant, A., and H. White, 1988, *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell, Oxford.

Geweke, J., 1975, "Employment turnover and wage dynamics in U.S. manufacturing," Ph.D. dissertation, University of Minnesota.

——— , 1978, "Temporal Aggregation in the Multiple Regression Model," *Econometrica*, 46, 643–661.

Ghysels, E., and D. Osborn, 2001, *The Econometric Analysis of Seasonal Time Series*. Cambridge University Press, Cambridge.

Ghysels, E., P. Santa-Clara, A. Sinko, and R. Valkanov, 2003, "MIDAS Regressions: Further Results and New Directions," Working paper, UNC and UCLA.

Ghysels, E., P. Santa-Clara, and R. Valkanov, 2002, "There is a risk-return tradeoff after all," *Journal of Financial Economics*, forthcoming.

——— , 2003, "Predicting volatility: getting the most out of return data sampled at different frequencies," Working paper, UNC and UCLA.

Greene, W., 2000, *Econometic Analysis*. Prentice Hall.

Hannan, E., 1963a, "Regression for Time Series," in *Proceedings of a Symposium on Time Series Analysis*, ed. by M. Rosenblatt. John Wiley.

——— , 1963b, "Regression for Time Series with Errors of Measurement," *Biometrika*, 50, 293–302.

——— , 1968, "Least-squares Efficiency for Vector Time Series," *Journal of the Royal Statistical Society. Series B*, 30, 490–498.

Hansen, L., and T. Sargent, 1983, "The Dimensionality of the Alliasing Problem with Rational Spectral Densities," *Econometrica*, 51, 377–387.

———, 1991a, "Identification of continuous time rational expectations from discrete data," in *Rational Expectations Econometrics*, ed. by L. Hansen, and T. Sargent, pp. 219–235. Westview Press, Boulder.

———, 1991b, "Two difficulties in interpreting vector autoregressions," in *Rational Expectations Econometrics*, ed. by L. Hansen, and T. Sargent, pp. 77–119. Westview Press, Boulder.

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

Judge, G., W. Griffith, R. Hill, H. Lutkepohl, and T. Lee, 1985, *The theory and Practice of Econometrics - Second Edition*. John Wiley & Sons.

McCrorie, J., 2000, "Deriving the exact discrete analog of a continuous time system," *Econometric Theory*, 16, 998–1015.

Phillips, A., 1959, "The estimation of parameters in systems of stochastic differential equations," *Biometrika*, 46, 67–76.

Phillips, P., 1972, "The Structural Estimation of a Stochastic Differential Equation System," *Econometrica*, 40, 1021–1041.

———, 1973, "The Problem of Identification in Finite Parameter Continuous Time Models," *Journal of Econometrics*, 1, 351–362.

———, 1974, "The Estimation of some Continuous Time Models," *Econometrica*, 42, 803–824.

Phillips, P., B. Guo, and Z. Xiao, 2002, "Efficient regression in time series partial linear models," Cowles Foundation Discussion paper No. 1363.

Robinson, P., 1977, "The construction and estimation of continuous time models and discrete approximations in econometrics," *Journal of Econometrics*, 6, 173–198.

———, 1988, "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931–954.

———, 1991, "Automatic frequency domain inference on semiparametric and nonparametric models," *Econometrica*, 59, 755–786.

Schwarz, G., 1978, "Estimating the dimension of a model," *Annals of Statistics*, 6, 461–464.

Sims, C., 1971, "Discrete Approximations to Continuous Time Distributed Lags in Econometrics," *Econometrica*, 39, 545–563.

——— , 1972, "The role of approximate prior restrictions in distributed lag estimation," *Journal of American Statistical Association*, 67, 169–175.

——— , 1974, "Distributed lags," in *Frontiers of Quantitative Economics II*, ed. by M. D. Intrilligator, and D. A. Kendrick. North-Holland, Amsterdam.

Stinchcombe, M., and H. White, 1998, "Consistent specification testing with nuisance parameters present only under the alternative," *Econometric Theory*, 14, 295–325.

Stock, J., and M. Watson, 2003, *Introduction to Econometrics*. Addison-Wesley.

Wallis, K., 1974, "Seasonal adjustment and relations between variables," *Journal of American Statistical Association*, 69, 618–626.

Wang, L., 2003, "On the Intertemporal Risk-Return Relation: A Bayesian Model Comparison Perspective," Working paper, Wharton.

Wooldridge, J., 1999, *Introductory Econometrics: A Modern Approach*. South-Western.

Xiao, Z., and P. Phillips, 1998, "Higher-order approximations for frequency domain time series regression," *Journal of Econometrics*, 86, 297–336.

A Regularity Conditions

It is worth recalling equation (3.3), namely:

$$\begin{aligned} y(t) &= b * x(t) + u(t) \\ &= \int_{-\infty}^{\infty} x(t-s)b(s)ds + u(t) \end{aligned}$$

where the errors are not necessarily i.i.d. In addition, the following technical conditions are assumed to hold:

Assumption A.1 *The continuous time processes $y(t)$, $x(t)$ and $u(t)$ are covariance stationary with spectral densities S_y, S_x, S_u and cross-spectrum S_{xy}.*

Assumption A.2 *To ensure identification of b in equation (3.3) rests on the assumption that the x process is, up to second moments, truly exogenous, i.e. $E[x(t)u(s)] = 0$, $\forall$ s and $t \in R$.*

So far, we did not distinguish single regressor and multiple regressor cases. In the main body of the paper we treated the single regressor case for ease of presentation. The following technical conditions cover the general multiple regression case.

Assumption A.3 *$b(s)$ in (3.3) is an N-dimensional vector of absolutely integrable functions of bounded total variation.*

Assumption A.4 *The eigenvalues of the spectral density matrix of $x(t)$ are strictly bounded away from zero on every finite frequency interval and that in the auxiliary regressions: $x_i(t) = \int_{-\infty}^{\infty} x_j(t-s)'b_{ij}(s)ds + \varepsilon_{ij}(t)$ all b_{ij} are ordinary absolutely integrable functions.*

In order to define the Hannan efficient estimators studied in section 3 we consider the spectral densities $F_m[S_x]$, $F_m[S_x]$ and cross-spectrum $F_m[S_{yx}]$, and define the estimators:

$$\hat{S}_Z^{(m)}(\lambda) = \frac{1}{2\pi m} \sum_{s=-k_m}^{k_m} K(\frac{s}{k_m}) \hat{R}_Z^{(m)}(s) \exp(-is\lambda) \tag{A.1}$$

for $Z^{(m)}$ equal to $X^{(m)}$ or $U^{(m)}$, where k_m is a bandwidth parameter, K is a kernel function and $\hat{R}_Z^{(m)}(s)$ is the autocovariance function of $Z^{(m)}$. Equation (A.1) can be straightforwardly modified to deal with the cross-spectrum $F_m[S_{yx}]$ as well. The following conditions are assumed to hold:

Assumption A.5 *The bandwidth $k_m = o(\sqrt{mT})$ and the kernel $K(x)$ is an even, bounded function for $x \in [-1, 1]$ with $k(0) = 1$ and $k(x) = 0$ for $x \notin [-1, 1]$.*

In addition, the following assumption is made regarding the error processes and autocovariances:

Assumption A.6 *The error processes $u_{t/m}^{(m)}$ and u_t in equations (3.10) and (3.11) respectively, are linear processes with absolute summable Wold decomposition moving average representations. Moreover, the autocovariances appearing in Assumption A.5, $R_Z^{(m)}(s)$ satisfy: $\sum_{s=-\infty}^{\infty}|s|\|R_Z^{(m)}(s)\|<\infty$.*

B Proof of Theorem 3.1

We start from the observation in Sims (1972) that the least squares estimator of $B^{(m)}$ in (3.4) minimizes the following criterion:

$$\int_{-\pi m}^{\pi m}|\tilde{B}^{(m)}(\omega)-\tilde{b}(\omega)|^2F_m[S_x](\omega) \tag{B.2}$$

When all processes are point-sampled, the residuals of the MIDAS regression (3.5) are skip-sampled versions of the residuals in (3.4). Consequently, the least squares estimator minimizes the criterion (B.2) subject to frequency folding:

$$\sum_{j=-\infty}^{\infty}\int_{-\pi}^{\pi}|\tilde{B}^{(m)}(\omega+2\pi j)-\tilde{b}(\omega+2\pi j)|^2F_m[S_x](\omega+2\pi j) \tag{B.3}$$

Since it is assumed that the spectral density matrix is strictly positive the minimizations in (B.2) and (B.3) yield the same minimand $\bar{B}^{(m)}$.

Another way of obtaining the equivalence between $B^{(m)}$ in (3.4) and $\bar{B}^{(m)}$ in (3.5) is to note that Sims (1971) shows the following relationship between continuous sampled b in (3.3) and the discretely sampled distributed lag estimator $B^{(m)}$ in (3.4):

$$\begin{aligned} B^{(m)} &= \int_{-\infty}^{\infty}r_x^{(m)}b(s)ds \\ r_x^{(m)}(t) &= \sum_{s=-\infty}^{\infty}R_{X^{(m)}}^{-*}(s)R_x(t-s) \end{aligned} \tag{B.4}$$

where $R_{X^{(m)}}(s)$ is the autocovariance function of $x(t)$ and $R_{X^{(m)}}^{-*}(s)$ is the inverse under convolution of $R_{X^{(m)}}(s)$.[22] From equation (B.4) Sims (1971) notes (p. 548) that $B^{(m)}$ is obtained from b by first smoothing b using the filter $R_x * R_{X^{(m)}}$ since another way of writing the above equation is $B^{(m)} = b(s) * R_x(t-s) * R_{X^{(m)}}^{-*}$. Clearly, this filter only involves $X^{(m)}$ and therefore the same relationship holds for $\bar{B}^{(m)}$.

The remainder of the proof follows Geweke (1978), following the same steps to show that (3.8) holds for MIDAS regressions. Likewise, to establish that (3.9) holds it is also straightforward to follow the proof in Geweke (1978).

[22] Equation (B.4) is derived for multivariate regressions in Geweke (1978).

C Proof of Theorem 3.2

In this section we present the proof of theorem 3.2. The result of the MIDAS regression estimator appearing in equation (3.13) is a straightforward application of the original result in Hannan (1963a) and Hannan (1963b) who imposes Assumptions A.1 through A.5 (with some slight modifications that are of no consequence here). In particular the MIDAS regression is viewed as a regression projection of Y onto $X^{(m)}$ in the context of the Hannan procedure. The result for the distributed lag estimator is similar to that appearing in Hannan (1963a) and Hannan (1963b), yet modified for the sampling at frequency $1/m$. The asymptotic distributions appearing in(3.14) and (3.15) follow as well.

We are interested in the efficiency comparison and therefore focus on the difference between the distribution of $\sqrt{T}(\hat{B}_M^{(m)} - b^{(m)})$ versus that of $\sqrt{mT}(\hat{B}_H^{(m)} - b^{(m)})$ (ignoring, as noted in the main body of the paper the potential differences in discretization biases with fixed m). First we re-scale the latter such that the asymptotic distribution is in terms of $\sqrt{T}$, yielding the covariance matrices:

$$2\frac{\pi}{m}\{\int_{-m\pi}^{m\pi} F_m[S_x(\omega)](F_m[S_u(\omega)])^{-1}d\omega\}^{-1})$$

and

$$2\pi\{\int_{-\pi}^{\pi} F_m[S_x(\omega)](F_1[S_u(\omega)])^{-1}d\omega\}^{-1})$$

Assuming that $F_m[S_u]$ is constant, i.e. $U^{(m)}$ is white noise, leads to the above being equal to:

$$2\frac{\pi}{m}\{\int_{-m\pi}^{m\pi} F_m[S_x(\omega)](\frac{\sigma_u^2}{2\pi m})^{-1}d\omega\}^{-1})$$

and

$$2\pi\{\int_{-\pi}^{\pi} F_m[S_x(\omega)](\frac{\sigma_u^2}{2\pi})^{-1}d\omega\}^{-1})$$

which are equivalent.