

# msreg: A command for consistent estimation of linear regression models using matched data

Masayuki Hirukawa  
Ryukoku University  
Kyoto, Japan  
hirukawa@econ.ryukoku.ac.jp

Di Liu  
StataCorp  
College Station, TX  
dliu@stata.com

Artem Prokhorov  
University of Sydney Business School  
Sydney, Australia, and  
CEBA, St. Petersburg State University  
St. Petersburg, Russia, and  
CIREQ, University of Montreal  
Montreal, Canada  
artem.prokhorov@sydney.edu.au

**Abstract.** Economists often use matched samples, especially when dealing with earning data where some observations are missing in one sample and need to be imputed from another sample. Hirukawa and Prokhorov (2018, *Journal of Econometrics* 203: 344–358) show that the ordinary least-squares estimator using matched samples is inconsistent and propose two consistent estimators. We describe a new command, `msreg`, that implements these two consistent estimators based on two samples. The estimators attain the parametric convergence rate if the number of continuous matching variables is no greater than four.

**Keywords:** `st0630`, `msreg`, bias correction, linear regression, matching estimation

## 1 Introduction

Matching-based imputation is common in economic datasets. For example, the U.S. Census uses a practice known as “hot-deck imputation”, which is implemented when census reports for nonresponders the values of important variables such as earnings and income are borrowed from responders with a few similar characteristics. In some surveys, the share of such imputed responses reaches 30%.

Hirukawa and Prokhorov (2018) were concerned with this widely used but often ignored practice. The concern was that users of such data in applied econometrics are often unaware that these are imputed, rather than actual, observations and that the resulting matching discrepancy leads to nonnegligible biases in the ordinary least-squares (OLS) estimator. They list many other settings, usually involving more than one dataset, where matching is unavoidable and needs to be accounted for.

The goal of this article is to facilitate the use of the consistent estimation approaches proposed by Hirukawa and Prokhorov (2018) in their numerous applications. Hirukawa

and Prokhorov (2018) derive the imputation bias analytically and propose two bias-corrected estimators. In this article, we introduce a new command, `msreg`, that implements both estimators in Stata.

Section 2 documents theoretical backgrounds for the `msreg` command. Section 3 discusses the `msreg` syntax and provides a numerical example. Section 4 contains a simulation study. Section 5 provides an empirical application by estimating the return to schooling as in Hirukawa and Prokhorov (2018).

## 2 Setup and estimators

### 2.1 Setup and assumptions

Suppose that we are interested in fitting a linear regression model,

$$y = \beta_0 + \mathbf{x}'_1\beta_1 + \mathbf{x}'_2\beta_2 + \mathbf{z}'\gamma + u = \mathbf{w}'\boldsymbol{\theta} + u \quad \mathbf{E}(u|\mathbf{w}) = 0$$

where  $\mathbf{x}_1 \in \mathbf{R}^{d_1}$ ,  $\mathbf{x}_2 \in \mathbf{R}^{d_2}$ ,  $\mathbf{z} \in \mathbf{R}^{d_z}$ ,  $\mathbf{w} = (1, \mathbf{x}'_1, \mathbf{x}'_2, \mathbf{z}')$ , and  $\boldsymbol{\theta} = (\beta_0, \beta_1', \beta_2', \gamma')'$ .

If we can observe all the variables in one sample, OLS is a consistent estimator for  $\boldsymbol{\theta}$ . However, in reality, we often encounter a situation where the variables are taken from two different samples. To be precise, we need more notation to distinguish between the two samples. The first sample is denoted by  $\mathbf{S}_1 = \{(y_i, \mathbf{x}_{1i}, \mathbf{z}_i)\}_{i=1}^n$ . The second sample is denoted by  $\mathbf{S}_2 = \{(\mathbf{x}_{2j}, \mathbf{z}_j)\}_{j=1}^m$ . For inference, we also denote  $d_3$  as the number of continuous common variables in  $\mathbf{z}$  hereafter, which is not always equal to  $d_z$ .

Estimation theories in Hirukawa and Prokhorov (2018) are built on a set of assumptions that are required for identification, consistency, and asymptotic normality of their estimators. Some of them are quite common. For example, assumption 2 imposes compactness of the support of continuous common variables. In our empirical analysis in section 5, `educ`, `feduc`, and `meduc` are such variables, and it is natural to think of their support as compact. On the other hand, there are more subtle assumptions in Hirukawa and Prokhorov (2018) that may or may not hold in a given application. Examples include a common joint distribution for  $\mathbf{S}_1$  and  $\mathbf{S}_2$  (assumption 1),  $\mathbf{E}(\boldsymbol{\eta}_1\boldsymbol{\eta}'_2) = 0$ , and strict nonlinearity in  $g_2(\cdot)$  [assumption 3(ii)], where  $\boldsymbol{\eta}_\ell := \mathbf{x}_\ell - g_\ell(\mathbf{z})$  and  $g_\ell(\mathbf{z}) := \mathbf{E}(\mathbf{x}_\ell|\mathbf{z})$  for  $\ell = 1, 2$ . It is difficult to test the validity of these assumptions because  $\mathbf{x}_1$  and  $\mathbf{x}_2$  belong to two distinct samples.

### 2.2 Nearest-neighbor matching

The matched sample can be constructed via the nearest-neighbor matching (NNM) using a vector of common variables  $\mathbf{z}$  across two samples. Note that  $\mathbf{z}$  must contain at least one continuous variable for valid inference; inclusion of discrete common variables with a finite number of support points (for example, binary variables) in  $\mathbf{z}$  does not affect the asymptotic results that will be stated shortly.

To specify the NNM, we need to first define a matrix norm to measure the distance between two vectors. For a vector  $\mathbf{x}$  and some symmetric and positive definite matrix  $\mathbf{A}$ , the vector norm is defined as  $\|\mathbf{x}\|_{\mathbf{A}} = (\mathbf{x}'\mathbf{A}\mathbf{x})^{1/2}$ . Following Abadie and Imbens (2011), we use the Mahalanobis metric  $\mathbf{A}_M = \left\{ (1/N) \sum_{i=1}^N (\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})' \right\}^{-1}$  and the normalized Euclidean metric  $\mathbf{A}_{NE} = \mathbf{diag}(\mathbf{A}_M^{-1})^{-1}$ , where  $N = n + m$  and  $\bar{\mathbf{z}} = (1/N) \sum_{i=1}^N \mathbf{z}_i$ .

Let  $j_k(i)$  be the index of the  $k$ th match in  $\mathbf{S}_2$  to the unit  $i$  in  $\mathbf{S}_1$ ; that is, for each  $i \in \{1, \dots, n\}$ ,  $j_k(i)$  satisfies

$$\sum_{j=1}^m \mathbf{1} \left\{ \|\mathbf{z}_j - \mathbf{z}_i\|_{\mathbf{A}} \leq \|\mathbf{z}_{j_k(i)} - \mathbf{z}_i\|_{\mathbf{A}} \right\} = k$$

In other words,  $\mathbf{z}_{j_k(i)}$  is the  $k$ th nearest neighbor in  $\mathbf{S}_2$  to the unit  $i$  in  $\mathbf{S}_1$ .

For each unit  $i$ , let  $\mathbf{J}_K(i) = \{j_i(1), \dots, j_i(K)\}$  denote the  $K$  matches from  $\mathbf{S}_2$ . The NNM-based matched sample is

$$\mathbf{S} = \left\{ (y_i, \mathbf{x}_{1i}, \mathbf{x}_{2j_1(i)}, \dots, \mathbf{x}_{2j_K(i)}, \mathbf{z}_i) \right\}_{i=1}^n$$

We also write  $\overline{\mathbf{x}_{2j(i)}} = (1/K) \sum_{j \in \mathbf{J}_K(i)} \mathbf{x}_{2j}$ .

For estimation, we use a transformation of the matched sample  $\mathbf{S}$ .

$$\mathbf{S}^* = \left\{ (y_i, \mathbf{x}_{1i}, \overline{\mathbf{x}_{2j(i)}}, \mathbf{z}_i) \right\}_{i=1}^n$$

In contrast with the original matched sample  $\mathbf{S}$ , we replace the individual matched variable  $\mathbf{x}_{2j_1(i)}, \dots, \mathbf{x}_{2j_K(i)}$  by its mean  $\overline{\mathbf{x}_{2j(i)}}$ .

Throughout, it is assumed that we estimate  $\boldsymbol{\theta}$  by regressing  $y_i$  on  $\mathbf{w}_{i,j(i)} = (1, \mathbf{x}'_{1i}, \overline{\mathbf{x}'_{2j(i)}}, \mathbf{z}_i)'$ .

### 2.3 Inconsistency of matched-sample OLS

We start by using an OLS estimator on the matched sample  $\mathbf{S}^*$ . The OLS estimator is

$$\widehat{\boldsymbol{\theta}}_{\text{MSOLS}} = \widehat{\mathbf{Q}}_W^{-1} \widehat{\mathbf{R}}_W$$

where  $\widehat{\mathbf{Q}}_W = (1/n) \sum_{i=1}^n \mathbf{w}_{i,j(i)} \mathbf{w}'_{i,j(i)}$  and  $\widehat{\mathbf{R}}_W = (1/n) \sum_{i=1}^n \mathbf{w}_{i,j(i)} y_i$ . It is referred to as the matched-sample OLS (MSOLS) estimator.

**Theorem 1** (Hirukawa and Prokhorov [2018], theorem 1). *Under some regularity conditions,*

$$\widehat{\boldsymbol{\theta}}_{\text{MSOLS}} = \mathbf{Q}_W^{-1} \mathbf{P}_W \boldsymbol{\theta} + O_p \left( m^{-1/d_3} \right) + O_p \left( n^{-1/2} \right)$$

where  $\mathbf{Q}_W = \mathbf{E}(\mathbf{w}_{i,j(i)} \mathbf{w}'_{i,j(i)})$ ,  $\mathbf{P}_W = \mathbf{Q}_W - (1/K) \boldsymbol{\Sigma}$ ,  $\boldsymbol{\Sigma}$  is a  $(d+1) \times (d+1)$  block-diagonal matrix of the form  $\boldsymbol{\Sigma} = \mathbf{diag}\{\mathbf{0}_{(d+1) \times (d+1)}, \boldsymbol{\Sigma}_2, \mathbf{0}_{d_z \times d_z}\}$ , and  $\boldsymbol{\Sigma}_2 = \mathbf{E}(\boldsymbol{\eta}_2 \boldsymbol{\eta}'_2)$ .

Theorem 1 implies that MSOLS is inconsistent in general. The inconsistency is attributed to correlation between the imputed regressor  $\bar{\mathbf{x}}_{2j(i)}$  and  $(1/K) \sum_{j \in \mathbf{J}_{K(i)}} \eta_{2j}$  in the composite error term  $\epsilon_{i,j(i)}$ . All asymptotic analyses in Hirukawa and Prokhorov (2018) are based on letting  $n$  and  $m$  diverge while keeping  $K$  fixed. It is in principle possible to restore consistency by letting  $K$  diverge at a rate slower than  $n$  and  $m$ . However, a fixed  $K$  is what researchers are likely to do in practice; Abadie and Imbens (2006) also adopt this setup. Moreover, the  $O_p(m^{-1/d_3})$  term corresponds to the second-order bias term  $\lambda_{i,j(i)}$  because of the matching discrepancy of Abadie and Imbens (2006). Observe that this term affects the convergence rate of  $\widehat{\boldsymbol{\theta}}_{\text{MSOLS}}$  to its probability limit; see remark 3 of Hirukawa and Prokhorov (2018) for more details.

## 2.4 One-step bias-corrected estimator

The source of the inconsistency of the MSOLS estimator is that  $\widehat{\mathbf{Q}}_W \xrightarrow{p} \mathbf{Q}_W$ , whereas  $\widehat{\mathbf{R}}_W \xrightarrow{p} \mathbf{P}_W \boldsymbol{\theta} = \{\widehat{\mathbf{Q}}_W - (1/K)\boldsymbol{\Sigma}\} \boldsymbol{\theta}$ . To eliminate the nonvanishing bias, we replace the denominator  $\widehat{\mathbf{Q}}_W$  by a consistent estimator of  $\mathbf{P}_W$  and leave the numerator  $\widehat{\mathbf{R}}_W$  unchanged. Because this bias correction has an indirect inference interpretation, Hirukawa and Prokhorov (2018) call this estimator the matched-sample indirect inference (MSII) estimator. Let  $\widehat{\mathbf{P}}_W$  be some consistent estimator of  $\mathbf{P}_W$ . Then, the MSII estimator is defined as

$$\widehat{\boldsymbol{\theta}}_{\text{MSII}} = \widehat{\mathbf{P}}_W^{-1} \widehat{\mathbf{R}}_W$$

To consistently estimate  $\mathbf{P}_W$ , we need consistent estimators for  $\mathbf{Q}_W$  and  $\boldsymbol{\Sigma}$ . Apparently,  $\widehat{\mathbf{Q}}_W$  is a natural estimator for  $\mathbf{Q}_W$ . Furthermore, it turns out that we can consistently estimate  $\boldsymbol{\Sigma}$  without nonparametric estimation of  $\mathbf{E}(\mathbf{x}_2|\mathbf{z})$ . To do so, we first reorder  $\mathbf{S}_2$  with respect to  $\mathbf{z}$  in ascending order.

1. Define  $\mathbf{z}_{(1)}$  as the observation that has the smallest first element; that is,  $(1) = \arg \min_{1 \leq j \leq m} \mathbf{z}_{j1}$ .
2. For  $j = 2, \dots, m$ , choose  $(j) = \arg \min_{j \neq (1), \dots, (j-1)} \|\mathbf{z}_j - \mathbf{z}_{(j-1)}\|$ , where the norm of a matrix  $\|\mathbf{A}\|$  is defined as  $\|\mathbf{A}\| = \{\text{tr}(\mathbf{A}'\mathbf{A})\}^{1/2}$ .

Given the reordered sample  $\mathbf{S}_2 = \{(\mathbf{x}_{2(j)}, \mathbf{z}_{(j)})\}_{j=1}^m$ ,  $\boldsymbol{\Sigma}_2$  can be consistently estimated by

$$\widehat{\boldsymbol{\Sigma}}_2 = \frac{1}{2(m-1)} \sum_{j=2}^m \Delta \mathbf{x}_{2(j)} \Delta \mathbf{x}'_{2(j)}$$

where  $\Delta \mathbf{x}_{2(j)} = \mathbf{x}_{2(j)} - \mathbf{x}_{2(j-1)}$ . This is known as the difference-based variance estimator; see von Neumann (1941), Yatchew (1997), and Horowitz and Spokoiny (2001) for references.

The estimator of  $\mathbf{P}_W$  is given by

$$\widehat{\mathbf{P}}_W = \widehat{\mathbf{Q}}_W - \frac{1}{K} \widehat{\boldsymbol{\Sigma}} = \widehat{\mathbf{Q}}_W - \frac{1}{K} \mathbf{diag} \left\{ \mathbf{0}_{(d_1+1) \times (d_1+1)}, \widehat{\boldsymbol{\Sigma}}_2, \mathbf{0}_{d_z \times d_z} \right\}$$

Theorem 2 below documents asymptotic normality of  $\widehat{\boldsymbol{\theta}}_{\text{MSII}}$ . The theorem applies only when the number of continuously distributed matching variables is so small that the second-order, matching discrepancy bias can be safely ignored. Observe that both the convergence rate of  $\widehat{\boldsymbol{\theta}}_{\text{MSII}}$  and its asymptotic variance depend on the divergence pattern of  $(n, m)$ .

**Theorem 2** (Hirukawa and Prokhorov [2018], Corollary 1). *Under some regularity conditions, as  $n, m \rightarrow \infty$ ,*

$$\left\{ \begin{array}{ll} \sqrt{n} \left( \widehat{\boldsymbol{\theta}}_{\text{MSII}} - \boldsymbol{\theta} \right) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_I) = N\left(\mathbf{0}, \mathbf{P}_W^{-1} \boldsymbol{\Omega} \mathbf{P}_W^{-1}\right) & \text{if } n/m \rightarrow \kappa \in (0, \infty) \\ & \text{and } d_3 = 1 \\ \sqrt{n} \left( \widehat{\boldsymbol{\theta}}_{\text{MSII}} - \boldsymbol{\theta} \right) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_{II}) = N\left(\mathbf{0}, \mathbf{P}_W^{-1} \boldsymbol{\Omega}_{11A} \mathbf{P}_W^{-1}\right) & \text{if } n/m \rightarrow 0 \\ & \text{and } d_3 = 1, 2 \\ \sqrt{m} \left( \widehat{\boldsymbol{\theta}}_{\text{MSII}} - \boldsymbol{\theta} \right) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_{III}) = N\left(\mathbf{0}, \mathbf{P}_W^{-1} \boldsymbol{\Omega}_{22} \mathbf{P}_W^{-1}\right) & \text{if } n/m \rightarrow \infty \\ & \text{and } d_3 = 1 \end{array} \right.$$

where the definitions of  $\boldsymbol{\Omega}$ ,  $\boldsymbol{\Omega}_{11A}$ , and  $\boldsymbol{\Omega}_{22}$  can be found in the appendix, along with their consistent estimates.

As demonstrated in this theorem and theorem 3 below, the bias-corrected estimators of Hirukawa and Prokhorov (2018) attain the parametric convergence rate only when the number of continuous common variables is four or less. It may be tempting to include as many continuous common variables as possible in the NNM. However, this results in slowing down the convergence rate, and we do not recommend it.

## 2.5 Two-step bias-corrected estimator

The one-step bias-corrected estimator can attain the parametric rate of convergence with at most two matching variables. To overcome this curse of dimensionality, we should eliminate the second-order bias  $\lambda_{i,j(i)}$ . The entire procedure is reminiscent of the fully modified least-squares estimation for cointegrating regressions by Phillips and Hansen (1990). In this sense, Hirukawa and Prokhorov (2018) call the estimator the fully modified MSII (MSII-FM) estimator.

Estimating  $\lambda_{i,j(i)}$  requires consistent estimates of  $\boldsymbol{\theta}$  and  $g_2(\cdot)$ . For  $\boldsymbol{\theta}$ , we can use the MSII estimate  $\widehat{\boldsymbol{\theta}}_{\text{MSII}}$ . For  $g_2(\cdot)$ , we use a nonparametric power-series estimation as in Abadie and Imbens (2011). Let  $v = (v_1, \dots, v_{d_z})$  be a multi-index of dimension

$d_z$ , which is a  $d_z$ -dimensional vector of nonnegative integers with  $|v| = \sum_{l=1}^{d_z} v_l$ . Also, denote  $z^v = \prod_{l=1}^{d_z} z_l^{v_l}$ . Consider a series  $\{v_Q\}_{Q=1}^{\infty}$  containing distinct vectors such that  $|v(Q)|$  is nondecreasing. Let  $p_Q(\mathbf{z}) = \mathbf{z}^{v(Q)}$  and  $p^Q(\mathbf{z}) = \{p_1(\mathbf{z}), \dots, p_Q(\mathbf{z})\}'$ . Then, a nonparametric series estimator of the regression function  $g_{2r}(\mathbf{z})$ ,  $r = 1, \dots, d_2$  is

$$\widehat{g}_{2r} = p^{Q(m)}(\mathbf{z})' \left\{ \sum_{j=1}^m p^{Q(m)}(\mathbf{z}_j) p^{Q(m)}(\mathbf{z}_j)' \right\}^{-} \sum_{j=1}^m p^{Q(m)}(\mathbf{z}_j) x_{2r,j}$$

where  $x_{2r,j}$  denotes the  $r$ th element of  $\mathbf{x}_{2j}$  in  $\mathbf{S}_2$ ,  $(\cdot)^{-}$  denotes the generalized inverse, and  $Q = Q(m)$  implies the dependence of  $Q$  on the sample size of  $\mathbf{S}_2$ .

The entire estimation procedure can be summarized in the following three steps:

1. Run MSII using the matched sample  $\mathbf{S}^*$  to obtain

$$\widehat{\boldsymbol{\theta}}_{\text{MSII}} = \left( \widehat{\boldsymbol{\beta}}_{II,0}', \widehat{\boldsymbol{\beta}}_{II,1}', \widehat{\boldsymbol{\beta}}_{II,2}', \widehat{\boldsymbol{\gamma}}_{II}' \right)'$$

2. Construct adjusted dependent variables  $\{y_i^+\}_{i=1}^n = \{y_i - \widehat{\lambda}_{i,j(i)}\}_{i=1}^n$ , where

$$\widehat{\lambda}_{i,j(i)} = \left( \widehat{g}_2(\mathbf{z}_i) - \frac{1}{K} \sum_{j \in \mathbf{J}_K(i)} \widehat{g}_2(\mathbf{z}_j) \right)' \widehat{\boldsymbol{\beta}}_{II,2}$$

3. Rerun MSII using the modified matched sample  $\mathbf{S}^+ = \{(y_i^+, \mathbf{x}_{1i}, \overline{\mathbf{x}}_{2j(i)}, \mathbf{z}_i)\}_{i=1}^n$  to obtain the final estimator

$$\widehat{\boldsymbol{\theta}}_{\text{MSII-FM}} = \widehat{\mathbf{P}}_W^{-1} \widehat{\mathbf{R}}_W^+ = \widehat{\mathbf{P}}_W^{-1} \frac{1}{n} \sum_{i=1}^n \mathbf{w}'_{i,j(i)} y_i^+$$

**Theorem 3** (Hirukawa and Prokhorov [2018], theorem 4). *Under some regularity conditions, as  $n, m \rightarrow \infty$ ,*

$$\left\{ \begin{array}{ll} \sqrt{n} \left( \widehat{\boldsymbol{\theta}}_{\text{MSII-FM}} - \boldsymbol{\theta} \right) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_I) & \text{if } n/m \rightarrow \kappa \in (0, \infty) \text{ and } d_3 = 2, 3 \\ \sqrt{n} \left( \widehat{\boldsymbol{\theta}}_{\text{MSII-FM}} - \boldsymbol{\theta} \right) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_{II}) & \text{if } n/m \rightarrow 0 \text{ and } d_3 = 3, 4 \\ \sqrt{m} \left( \widehat{\boldsymbol{\theta}}_{\text{MSII-FM}} - \boldsymbol{\theta} \right) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_{III}) & \text{if } n/m \rightarrow \infty \text{ and } d_3 = 2, 3 \end{array} \right.$$

where the definitions of  $\mathbf{V}_I$ ,  $\mathbf{V}_{II}$ , and  $\mathbf{V}_{III}$  are the same as in theorem 2.

In practice, the standard errors (SEs) resulting from each of the three cases may be quite different. The relative magnitudes of  $n$  and  $m$  determine which case applies. Hirukawa and Prokhorov (2018) did not provide any generic comparisons for the variance matrices. Besides the scaling factor, the differences can be attributed to the specific features of the datasets and model specification. In borderline cases, it is advisable to use larger SEs for conservative inference.

## 3 The msreg command

### 3.1 Syntax

msreg has the following syntax.

```
msreg depvar [varlist_X1] (varlist_X2 = varlist_Z) using filename [if] [in]
    [, vce(vce_spec) estimator(est_spec) nneighbor(#) metric(metric_spec)
    order(#) noconstant level(#) display_options coeflegend]
```

### 3.2 Options

`vce(vce_spec)` specifies the type of variance–covariance matrix used in computation. `vce_spec` can be one of `vi`, `vii`, or `viii`. The default is `vce(vi)`. The definition of `vi`, `vii`, and `viii` can be found in theorem 2.

`estimator(est_spec)` specifies the type of estimator. `est_spec` can be either `onestep` or `twostep`. `onestep` specifies to use the one-step bias-corrected estimator. `twostep` specifies to use the two-step bias-corrected estimator. The default is `estimator(twostep)`.<sup>1</sup>

`nneighbor(#)` specifies the number of matches per observation. The default is `nneighbor(1)`. The maximum allowed number of matches is 10. Each observation is matched with the mean of the specified number of observations from the other dataset.

`metric(metric_spec)` specifies the distance matrix that is used as the weight matrix in a quadratic form that transforms the multiple distances into a single distance measure. `metric_spec` can be either `mahalanobis` or `euclidean`. `metric(mahalanobis)` specifies to use the inverse of the sample covariance matrix of matching variables, which is the default. `metric(euclidean)` specifies to use the inverse of only diagonal elements of the sample covariance matrix of matching variables.

`order(#)` specifies the order of polynomials in the power-series approximation for MSIFM. The default is `order(2)`. The maximum allowed number of order is 5.

`noconstant` suppresses the constant term.

`level(#)` specifies the level of significance for the output table.

`display_options`: `noci`, `nopvalues`, `noomitted`, `vsquish`, `noemptycells`, `baselevels`, `allbaselevels`, `nofvlabel`, `fvwrap(#)`, `fvwrapon(style)`, `cformat(%fmt)`, `pformat(%fmt)`, `sformat(%fmt)`, and `nolstretch`; see [R] **Estimation options**.

`coeflegend` specifies that the legend of the coefficients and how to specify them in an expression be displayed rather than displaying the statistics for the coefficients.

---

1. We intentionally leave the option `estimator(msols)` undocumented. MSOLS is inconsistent and we include it in the article for simulations only. We do not recommend using the option `estimator(msols)`.

### 3.3 Stored results

`msreg` stores the following in `e()`:

#### Scalars

<code>e(N_1)</code>	number of observations in the first sample
<code>e(N_2)</code>	number of observations in the second sample
<code>e(nneighbor)</code>	number of nearest neighbors matched
<code>e(chi2)</code>	chi-squared
<code>e(p)</code>	<i>p</i> -value for test of variables
<code>e(df_m)</code>	degree of freedom in the model
<code>e(order)</code>	order of polynomial in the power-series estimation

#### Macros

<code>e(cmd)</code>	<code>msreg</code>
<code>e(title)</code>	title in estimation output
<code>e(chi2type)</code>	Wald
<code>e(vce)</code>	<i>vce_spec</i> specified in <code>vce()</code>
<code>e(vcetype)</code>	title used to label Std. Err.
<code>e(properties)</code>	<code>b V</code>
<code>e(footnote)</code>	footnote displayed under the output table
<code>e(common)</code>	common variables that exist in both samples
<code>e(matched)</code>	matched variables
<code>e(metric)</code>	type of distance matrix
<code>e(estimator)</code>	type of estimator

#### Matrices

<code>e(b)</code>	coefficient vector
<code>e(V)</code>	variance–covariance matrix of the estimators

#### Functions

<code>e(sample)</code>	marks estimation sample
------------------------	-------------------------

### 3.4 A numerical example

We illustrate the use of `msreg` with a numerical example.

For illustration, we simulate two datasets: `s1.dta` and `s2.dta`. The first sample, `s1.dta`, contains the dependent variable `y` and some independent variables, `x11`, `x12`, `z1`, and `z2`. The second sample, `s2.dta`, contains some other dependent variables `x21`, `x22`, `z1`, and `z2`. Notice that variables `z1` and `z2` exist in both samples. In contrast, the variables `x21` and `x22` exist only in the second sample, `s2.dta`. The data-generating process is described in section 4.

We want to fit the following regression model:

$$y = \beta_0 + \beta_{11}x_{11} + \beta_{12}x_{12} + \beta_{21}x_{21} + \beta_{22}x_{22} + \gamma_1z_1 + \gamma_2z_2 + u \quad (1)$$

The true values of all the coefficients are set to 1.

Apparently, we cannot estimate (1) using just the first sample, `s1.dta`, because the variables `x21` and `x22` are missing in this dataset. Instead, we want to use the common variables that exist in both samples, that is, `z1` and `z2`, to construct matched variables `x22` and `x21` from the second sample, `s2.dta`.

We now use `msreg` to estimate the coefficients in (1). We can use the default two-step bias-corrected estimator and the default `vi` type variance if we assume  $n/m$  converges to a nonzero constant.

```
. use s1
. msreg y x11 x12 (x21 x22 = z1 z2) using s2, nneighbor(2) order(3)
Matched sample regression      Number of obs in sample 1 =      250
                               Number of obs in sample 2 =      750
                               Number of matches =              2
                               Order of power-series =            3
Estimator      : twostep      Wald chi2(6) =      5060.15
Distance metric : mahalanobis Prob > chi2 =      0.0000
```

	VI					
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
x11	.8993908	.1602669	5.61	0.000	.5852736	1.213508
x12	.9877756	.1594351	6.20	0.000	.6752885	1.300263
x21	1.103668	.0568935	19.40	0.000	.9921588	1.215177
x22	1.063374	.161514	6.58	0.000	.7468123	1.379936
z1	1.105623	.2745877	4.03	0.000	.5674414	1.643805
z2	.7304006	.2866126	2.55	0.011	.1686502	1.292151
_cons	.7213349	.2237863	3.22	0.001	.2827219	1.159948

Common variables : z1 z2

Matched variables: x21 x22

Note: VI Std. Err. assumes that sample size ratio between sample 1 and sample 2 converges to a nonzero constant and there are at most three continuous common variables.

Here are some comments on the syntax.

- The `(x21 x22 = z1 z2)` specifies that the variables `x21` and `x22` are the variables to be matched, and the variables `z1` and `z2` are the common variables that exist in both samples.
- using `s2` specifies that the variables `x21` and `x22` come from `s2.dta`.
- We use the default two-step bias-corrected estimator.
- The default option `vce(vi)` specifies to use the `vi`-type variance matrix as specified in theorem 2 because we assume the sample-size ratio between the two samples converges to a nonzero constant and there are only two continuous common variables used for matching. Actually, the note states a similar explanation about `vce(vi)`.
- Option `nneighbor(2)` specifies to pick out 2 matches via the NNM.
- Option `order(3)` specifies to fit a third-order polynomial in the power-series approximation for MSII-FM.
- The output shows the point estimates of coefficients and their SEs, and they can be interpreted as in a regular linear regression framework.

## 4 A simulation study

### 4.1 Simulation design

We conduct a Monte Carlo simulation study for two purposes: first, we want to see the finite-sample performance of MSII-FM in contrast to MSOLS; second, the simulation results can serve as a verification of the numerical implementation of our command `msreg`. The simulation study replicates that of Hirukawa and Prokhorov (2018).

The model considered throughout is

$$y = \beta_0 + \mathbf{x}'_1\boldsymbol{\beta}_1 + \mathbf{x}'_2\boldsymbol{\beta}_2 + \mathbf{z}'\boldsymbol{\gamma} + u \quad (2)$$

where  $\mathbf{x}_1 = (x_{11}, x_{12})'$ ,  $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12})' \in \mathbf{R}^2$ ,  $\mathbf{x}_2 = (x_{21}, x_{22})'$ ,  $\boldsymbol{\beta}_2 = (\beta_{21}, \beta_{22})' \in \mathbf{R}^2$ , and  $\mathbf{z} = (z_1, \dots, z_{d_3})'$ ,  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{d_3})' \in \mathbf{R}^{d_3}$  for  $d_3 = 1, 2, 3$ . We assume that two samples, namely,  $\mathbf{S}_1 = \{(y_i, \mathbf{x}_{1i}, \mathbf{z}_i)\}_{i=1}^n$  and  $\mathbf{S}_2 = \{(\mathbf{x}_{2j}, \mathbf{z}_j)\}_{j=1}^m$  are observable in practice.

Here is how to generate the data. First,  $\mathbf{z}^* = (z_1^*, z_2^*, z_3^*)$  is generated by

$$\mathbf{z}^* \stackrel{\text{i.i.d.}}{\sim} N \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{2} & 1 & \sqrt{2}/\sqrt{3} \\ 1/\sqrt{3} & \sqrt{2}/\sqrt{3} & 1 \end{bmatrix} \right)$$

Each  $z_p^*$  ( $p = 1, 2, 3$ ) is transformed to  $z_p = 4\Phi(z_p^*) - 2$ , where  $\Phi(\cdot)$  is the cdf of  $N(0, 1)$ . Notice that  $z_p$  are mutually correlated  $U[-2, 2]$  random variables. For a given  $d_3$ , the  $z_p$  ( $p \leq d_3$ ) are used as matching variables.

Second,  $\mathbf{x}_1 = (x_{11}, x_{12})'$  is generated by  $x_{1q} = \sum_{p=1}^{d_3} z_p + \eta_q$  ( $q = 1, 2$ ), where  $\eta_q \sim N(0, 1)$ . Third,  $\mathbf{x}_2 = (x_{21}, x_{22})'$  is generated by  $x_{2r} = \sum_{p=1}^{d_3} g_{2r}(z_p) + \eta_{2r}$  ( $r = 1, 2$ ) for some nonlinear function  $g_{2r}(\cdot)$ , where  $\eta_{2r} \sim N(0, 1)$ . Specifically,  $g_{21}(z) = z + (5/\tau)\phi(z/\tau)$ ,  $\tau = 0.25$ , where  $\phi(\cdot)$  is the pdf of  $N(0, 1)$ , and  $g_{22}(z) = 4\sqrt{|z/2|(1 - |z/2|)} \sin\{2\pi(1 + \epsilon)/(|z/2| + \epsilon)\}$ , with  $\epsilon = 0.05$ .

Finally,  $y$  is generated by setting all coefficients equal to 1 with  $u \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ . The sample sizes are set to  $(n, m) = (1000, 1000)$ . The number of replications is 1,000.

We focus on the finite-sample properties of estimators of  $\beta_{22}$  and  $\gamma_1$ . For each estimator, the following performance measures are computed: i) Bias ( $1 - \text{Mean}$ ), where Mean is the simulation average of the parameter estimate; ii) standard deviation (SD) (simulation SD of the parameter estimate); iii)  $\overline{\text{SE}}$  (simulation average of the SE); and iv) Rej. rate (rejection rate for the test of parameter estimate equal to its true value 1 against the nominal 5% level of significance).

For  $d_3 = 1, 2, 3$ , we estimate the coefficients in (2) using both MSOLS and MSII-FM. For MSOLS, the number of matches is  $K = 1, 2, 4, 8$ . For MSII-FM, the number of matches  $K$  is fixed at 1, and orders of polynomials in the power-series approximation are 2, 3, and 4. For a more complete simulation study, see section 4 of Hirukawa and Prokhorov (2018).

## 4.2 Results

The simulation results are summarized in tables 1 and 2 for MSOLS and MSII-FM, respectively.

Table 1 shows that regardless of the number of matches, there is a big bias of  $\beta_{22}$ , and there is a large rejection rate, which indicate inconsistency of MSOLS as implied in theorem 1.

Table 1. Monte Carlo results for MSOLS

$K$	$\beta_{22}$				$\gamma_1$			
	1	2	4	8	1	2	4	8
$d_3 = 1$								
Bias	0.4486	0.2866	0.1634	0.0773	-0.1680	-0.0919	-0.0442	-0.0216
SD	0.0426	0.0455	0.0474	0.0492	0.1122	0.1052	0.1005	0.0980
$\overline{SE}$	0.0507	0.0523	0.0546	0.0603	0.1174	0.1141	0.1119	0.1109
Rej. rate	1.0000	1.0000	0.9110	0.3800	0.3210	0.1550	0.1070	0.0910
$d_3 = 2$								
Bias	0.5280	0.3724	0.2239	0.0828	-0.1289	-0.0723	-0.0372	0.0093
SD	0.0408	0.0460	0.0523	0.0619	0.1468	0.1398	0.1369	0.1379
$\overline{SE}$	0.0462	0.0529	0.0627	0.0754	0.1548	0.1502	0.1534	0.1667
Rej. rate	1.0000	1.0000	0.9660	0.3110	0.1640	0.1030	0.0880	0.1100
$d_3 = 3$								
Bias	0.7532	0.6149	0.4472	0.2305	-0.2206	-0.1306	-0.0475	0.0281
SD	0.0472	0.0575	0.0707	0.0893	0.1993	0.1906	0.1881	0.1922
$\overline{SE}$	0.0514	0.0648	0.0794	0.1082	0.2015	0.1990	0.2078	0.2270
Rej. rate	1.0000	1.0000	1.0000	0.6860	0.1980	0.1210	0.0860	0.0920

Table 2 shows that 1) the bias is small; that is, the mean of the point estimates is very close to its true value; 2) the SD of the point estimate is very close to the mean of the SEs; and 3) the overall rejection rate is close to the nominal 5% level. Notice that for the case  $d_3 = 2$ , the rejection rate is a little bit off for  $\beta_{22}$ . However, based on results for larger samples, it seems that the over-rejection rate is due to the finite-sample bias of MSII-FM (reported in the Supplement of Hirukawa and Prokhorov [2018]). The simulation result shows that MSII-FM performs well in a finite sample as predicted by theorem 3 and that it also numerically verifies the implementation of `msreg`.

Table 2. Simulation results for MSII-FM

Order	$\beta_{22}$			$\gamma_1$		
	2	3	4	2	3	4
$d_3 = 1$						
Bias	-0.0305	-0.0305	-0.0323	0.0110	0.0110	0.0119
SD	0.1049	0.1049	0.1047	0.1257	0.1259	0.1258
SE	0.1130	0.1148	0.1149	0.1353	0.1361	0.1359
Rej. rate	0.0620	0.0610	0.0640	0.0730	0.0730	0.0720
$d_3 = 2$						
Bias	-0.1740	-0.1735	-0.1641	0.0318	0.0382	0.0401
SD	0.1539	0.1541	0.1540	0.1750	0.1754	0.1765
SE	0.1637	0.1636	0.1643	0.1912	0.1941	0.1930
Rej. rate	0.1400	0.1380	0.1270	0.0820	0.0840	0.0760
$d_3 = 3$						
Bias	-0.0948	-0.0904	-0.0866	0.0372	0.0408	0.0526
SD	0.2884	0.2925	0.2904	0.2481	0.2499	0.2495
SE	0.3041	0.3152	0.3132	0.2680	0.2749	0.2736
Rej. rate	0.0370	0.0350	0.0370	0.0610	0.0690	0.0650

## 5 An empirical application: Return to schooling

We now apply `msreg` to a version of Mincer's (1974) wage equation. We consider the following wage regression,

$$\begin{aligned} \log(\text{wage}) = & \beta_0 + \beta_1 \text{expr} + \beta_2 \text{expr}^2 + \beta_3 \text{kww} + \beta_4 \text{educ} + \beta_5 \text{feduc} + \beta_6 \text{meduc} \\ & + \beta_7 \text{smsa} + \beta_8 \text{south} + \beta_9 \text{black} + u \end{aligned} \quad (3)$$

where `expr` is years of experience, `educ` is years of education, `kww` is Knowledge of World of Work test score, `feduc` and `meduc` are years of father's and mother's education, and `black`, `smsa`, and `south` are dummy variables to indicate whether an individual is black, lives in an urban area, and lives in the south, respectively.

We can estimate (3) using only `card.dta` from Card (1995), as in the benchmark OLS result below.

```
. use card
. regress lnwage exper c.exper#c.exper
> kww educ feduc meduc i.smsa i.south i.black
```

Source	SS	df	MS	Number of obs	=	2,191
Model	114.487495	9	12.7208327	F(9, 2181)	=	89.96
Residual	308.394118	2,181	.141400329	Prob > F	=	0.0000
				R-squared	=	0.2707
				Adj R-squared	=	0.2677
Total	422.881613	2,190	.193096627	Root MSE	=	.37603

  

lnwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
exper	.0786594	.0082641	9.52	0.000	.062453 .0948658
c.exper#					
c.exper	-.0021546	.0004049	-5.32	0.000	-.0029486 -.0013607
kww	.0055563	.0013232	4.20	0.000	.0029615 .0081511
educ	.061174	.0051257	11.93	0.000	.0511222 .0712258
feduc	-.0018021	.0030073	-0.60	0.549	-.0076996 .0040955
meduc	.0070962	.0035015	2.03	0.043	.0002295 .013963
1.smsa	.1516693	.0186575	8.13	0.000	.1150809 .1882577
1.south	-.1111491	.0177506	-6.26	0.000	-.1459589 -.0763393
1.black	-.1320785	.0254406	-5.19	0.000	-.1819688 -.0821882
_cons	4.68611	.0815838	57.44	0.000	4.52612 4.8461

```
. estimate store ols
```

The estimation result is stored as `ols`.

Nonetheless, we pretend that the variable `kww` is missing in this dataset. In accordance with this scenario, we use yet another dataset, `wage2.dta`, from Blackburn and Neumark (1992). The dataset contains six variables—`educ`, `feduc`, `meduc`, `smsa`, `south`, and `black`—other than `kww`. All six variables are used as matching variables to impute the missing `kww`, where `educ`, `feduc`, and `meduc` are assumed to be continuous. Our aim is to see how the estimation result of (3) changes if `kww` is imputed from `wage2.dta`.

```

. msreg lnwage exper c.exper#c.exper
>      (kww = educ feduc meduc i.smsa i.south i.black)
>      using wage2, order(3)
Matched sample regression      Number of obs in sample 1 =      2,191
                                Number of obs in sample 2 =      470
                                Number of matches =      1
                                Order of power-series =      3
Estimator      : twostep      Wald chi2(9) =      785.11
Distance metric : mahalanobis Prob > chi2 =      0.0000

```

	VI					
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
exper	.0840744	.0082734	10.16	0.000	.0678589	.1002898
c.exper#						
c.exper	-.0021476	.0004202	-5.11	0.000	-.0029713	-.001324
kww	.0020755	.0052253	0.40	0.691	-.0081658	.0123168
educ	.0680349	.0079751	8.53	0.000	.0524041	.0836658
feduc	-.0005824	.003143	-0.19	0.853	-.0067427	.0055778
meduc	.0070538	.0039251	1.80	0.072	-.0006393	.0147469
1.smsa	.1557546	.0186934	8.33	0.000	.1191161	.192393
1.south	-.1114039	.0178276	-6.25	0.000	-.1463453	-.0764626
1.black	-.1517787	.0354906	-4.28	0.000	-.221339	-.0822183
_cons	4.654032	.1187133	39.20	0.000	4.421359	4.886706

```

Common variables : educ feduc meduc 0b.smsa 1.smsa 0b.south 1.south 0b.black
                  1.black

```

```

Matched variables: kww

```

```

Note: VI Std. Err. assumes that sample size ratio between sample 1 and sample
      2 converges to a nonzero constant and there are at most three continuous
      common variables.

```

```

. estimate store twostep_vi

```

We use the default `vi` type covariance estimation assuming the sample-size ratio between the two datasets converge to a nonzero constant. We use a third-order polynomial in power-series estimation to remove the second-order bias. The estimation result is stored as `twostep_vi`.

We can now compare these two estimation results.

```
. estimate table ols twostep_vi, b star
```

Variable	ols	twostep_vi
exper	.0786594***	.08407435***
c.exper#		
c.exper	-.00215464***	-.00214765***
kww	.00555626***	.00207551
educ	.06117396***	.06803494***
feduc	-.00180207	-.00058241
meduc	.00709624*	.00705379
smsa		
1	.15166928***	.15575459***
south		
1	-.11114906***	-.11140395***
black		
1	-.13207849***	-.15177868***
_cons	4.6861104***	4.6540324***

Legend: \* p<0.05; \*\* p<0.01; \*\*\* p<0.001

The first column shows the benchmark OLS results. The signs of `exper`, `exper2`, `kww`, and `educ` are as expected, and the estimates are significant at the 5% level.

The second column shows the results from the two-step bias-corrected estimator with the default `vi`-type covariance. All the point estimates have the same sign as in the OLS benchmark. However, the coefficient on `kww` is insignificant because of a large SE.

## 6 Conclusion

In this article, we described a new command, `msreg`, that implements two estimators proposed in Hirukawa and Prokhorov (2018). The command allows users to obtain consistent estimators of linear regression models after imputing missing regressors via the NNM. We illustrated the use of `msreg` through a numerical example and an empirical application.

## 7 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 21-1
. net install st0630      (to install program files, if available)
. net get st0630         (to install ancillary files, if available)
```

## 8 References

- Abadie, A., and G. W. Imbens. 2006. Large sample properties of matching estimators for average treatment effects. *Econometrica* 74: 235–267. <https://doi.org/10.1111/j.1468-0262.2006.00655.x>.
- . 2011. Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics* 29: 1–11. <https://doi.org/10.1198/jbes.2009.07333>.
- Blackburn, M., and D. Neumark. 1992. Unobserved ability, efficiency wages, and interindustry wage differentials. *Quarterly Journal of Economics* 107: 1421–1436. <https://doi.org/10.2307/2118394>.
- Card, D. E. 1995. Using geographic variation in college proximity to estimate the return to schooling. In *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, ed. L. N. Christofides, E. K. Grant, and R. Swidinsky, 201–222. Toronto, Canada: University of Toronto Press.
- Hirukawa, M., and A. Prokhorov. 2018. Consistent estimation of linear regression models using matched data. *Journal of Econometrics* 203: 344–358. <https://doi.org/10.1016/j.jeconom.2017.07.006>.
- Horowitz, J. L., and V. G. Spokoiny. 2001. An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* 69: 599–631. <https://doi.org/10.1111/1468-0262.00207>.
- Mincer, J. 1974. *Schooling, Experience, and Earnings*. New York: National Bureau of Economic Research.
- von Neumann, J. 1941. Distribution of the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics* 12: 367–395. <https://doi.org/10.1214/aoms/1177731677>.
- Phillips, P. C. B., and B. E. Hansen. 1990. Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57: 99–125. <https://doi.org/10.2307/2297545>.
- Yatchew, A. 1997. An elementary estimator of the partial linear model. *Economic Letters* 57: 135–143. [https://doi.org/10.1016/S0165-1765\(97\)00218-8](https://doi.org/10.1016/S0165-1765(97)00218-8).

### About the authors

Masayuki Hirukawa is a professor in the Faculty of Economics at Ryukoku University in Japan. Research for this article was supported by a grant from the Japan Society for the Promotion of Science (Project No. 19K01595).

Di Liu is a senior econometrician at StataCorp.

Artem Prokhorov is a professor in the Discipline of Business Analytics at the University of Sydney, Sydney, Australia, and a research fellow at CEBA, St. Petersburg State University, St. Petersburg, Russia, and CIREQ, University of Montreal, Montreal, Canada. Research for this article was supported by a grant from the Russian Science Foundation (Project No. 20-18-00365).

## A Appendix

Theorems 2 and 3 give the asymptotic distributions of  $\widehat{\boldsymbol{\theta}}_{\text{MSII}}$  and  $\widehat{\boldsymbol{\theta}}_{\text{MSII-FM}}$ , respectively. The covariance matrices  $\mathbf{V}_I$ ,  $\mathbf{V}_{II}$ , and  $\mathbf{V}_{III}$  depend on the definitions of  $\boldsymbol{\Omega}$ ,  $\boldsymbol{\Omega}_{11A}$ , and  $\boldsymbol{\Omega}_{22}$ . We define  $\boldsymbol{\Omega}$ ,  $\boldsymbol{\Omega}_{11A}$ , and  $\boldsymbol{\Omega}_{22}$  as follows:

$$\begin{aligned}\boldsymbol{\Omega} &= \boldsymbol{\Omega}_{11A} + \kappa \left[ (\boldsymbol{\beta}'_2 \boldsymbol{\Sigma}_2 \boldsymbol{\beta}_2) \mathbf{E}(\mathbf{w}) \mathbf{E}(\mathbf{w})' \right. \\ &\quad \left. + \frac{1}{K^2} \mathbf{diag} \left\{ \mathbf{0}_{(d_1+1) \times (d_1+1)}, (\boldsymbol{\beta}'_2 \boldsymbol{\Sigma}_2 \boldsymbol{\beta}_2) \mathbf{V}_{g_2} + \frac{1}{2} \boldsymbol{\Psi}, \mathbf{0}_{d_z \times d_z} \right\} \right] \\ \boldsymbol{\Omega}_{11A} &= \mathbf{E} \left\{ \left( \mathbf{w}_{i,j(i)} \epsilon_{i,j(i)} + \frac{1}{K} \boldsymbol{\Sigma} \boldsymbol{\theta} \right) \left( \mathbf{w}_{i,j(i)} \epsilon_{i,j(i)} + \frac{1}{K} \boldsymbol{\Sigma} \boldsymbol{\theta} \right)' \right\} \\ \boldsymbol{\Omega}_{22} &= \frac{1}{K^2} \mathbf{diag} \left\{ \mathbf{0}_{(d_1+1) \times (d_1+1)}, \boldsymbol{\Xi} + \frac{1}{2} \boldsymbol{\Psi}, \mathbf{0}_{d_z \times d_z} \right\} \\ \mathbf{V}_{g_2} &= \text{Var} \{g_2(\mathbf{z})\} \\ \boldsymbol{\Xi} &= \mathbf{E} \{ (\boldsymbol{\eta}_2 \boldsymbol{\eta}'_2 - \boldsymbol{\Sigma}_2) \boldsymbol{\beta}_2 \boldsymbol{\beta}'_2 (\boldsymbol{\eta}_2 \boldsymbol{\eta}'_2 - \boldsymbol{\Sigma}_2) \} \\ \boldsymbol{\Psi} &= (\boldsymbol{\beta}'_2 \boldsymbol{\Sigma}_2 \boldsymbol{\beta}_2) \boldsymbol{\Sigma}_2 + \boldsymbol{\Sigma}_2 \boldsymbol{\beta}_2 \boldsymbol{\beta}'_2 \boldsymbol{\Sigma}_2\end{aligned}$$

We present consistent estimators of  $\boldsymbol{\Omega}_{11A}$ ,  $\boldsymbol{\Omega}_{22}$ , and  $\boldsymbol{\Omega}$  for MSII below. Because MSII-FM is first-order asymptotically equivalent to MSII as documented in theorem 3, simply replacing  $\widehat{\boldsymbol{\theta}}_{\text{MSII}}$  in  $\widehat{\boldsymbol{\Omega}}_{11A}$ ,  $\widehat{\boldsymbol{\Omega}}_{22}$ , and  $\widehat{\boldsymbol{\Omega}}$  with  $\widehat{\boldsymbol{\theta}}_{\text{MSII-FM}}$  yields the estimators for MSII-FM,

$$\begin{aligned}
\widehat{\Omega}_{11A} &= \frac{1}{n} \sum_{i=1}^n \left( \mathbf{w}_{i,j(i)} \widehat{\epsilon}_{i,j(i)} + \frac{1}{K} \widehat{\Sigma} \widehat{\boldsymbol{\theta}}_{\text{MSII}} \right) \left( \mathbf{w}_{i,j(i)} \widehat{\epsilon}_{i,j(i)} + \frac{1}{K} \widehat{\Sigma} \widehat{\boldsymbol{\theta}}_{\text{MSII}} \right)' \\
\widehat{\Omega}_{22} &= \frac{1}{K^2} \text{diag} \left\{ \mathbf{0}_{(d_1+1) \times (d_1+1)}, \widehat{\Gamma}(-1) + \widehat{\Gamma}(0) + \widehat{\Gamma}(1), \mathbf{0}_{d_z \times d_z} \right\} \\
\widehat{\Omega} &= \widehat{\Omega}_{11A} + \frac{n}{m} \left( \left( \widehat{\boldsymbol{\beta}}_{2,\text{MSII}}' \widehat{\boldsymbol{\Sigma}}_2 \widehat{\boldsymbol{\beta}}_{2,\text{MSII}} \right) \overline{\mathbf{w}} \overline{\mathbf{w}}' \right. \\
&\quad \left. + \frac{1}{K^2} \text{diag} \left[ \mathbf{0}_{(d_1+1) \times (d_1+1)}, \widehat{\boldsymbol{\beta}}_{2,\text{MSII}}' \widehat{\boldsymbol{\Sigma}}_2 \widehat{\boldsymbol{\beta}}_{2,\text{MSII}} \widehat{\mathbf{V}}_{g_2} \right. \right. \\
&\quad \left. \left. + \widehat{\Gamma}(0) - \left\{ \widehat{\Gamma}(-1) + \widehat{\Gamma}(1) \right\}, \mathbf{0}_{d_z \times d_z} \right] \right)
\end{aligned}$$

where  $\widehat{\epsilon}_{i,j(i)} = y_i - \mathbf{w}'_{i,j(i)} \widehat{\boldsymbol{\theta}}_{\text{MSII}}$ ,  $\widehat{\boldsymbol{\beta}}_{2,II}$  is the MSII estimate of  $\boldsymbol{\beta}_2$ , and  $\widehat{\Gamma}(l)$  is the  $l$ th sample autocovariance of  $\left\{ \Delta \mathbf{x}_{2j} \Delta \mathbf{x}'_{2j} / 2 - \widehat{\boldsymbol{\Sigma}}_2 \right\}$ ; that is,

$$\begin{aligned}
\widehat{\Gamma}(l) &= \frac{1}{m-1} \sum_{j=\max\{2,2+l\}}^{\min\{m,m+l\}} \left( \frac{\Delta \mathbf{x}_{2j} \Delta \mathbf{x}'_{2j}}{2} - \widehat{\boldsymbol{\Sigma}}_2 \right) \widehat{\boldsymbol{\beta}}_{2,\text{MSII}} \widehat{\boldsymbol{\beta}}_{2,\text{MSII}}' \left( \frac{\Delta \mathbf{x}_{2j-l} \Delta \mathbf{x}'_{2j-l}}{2} - \widehat{\boldsymbol{\Sigma}}_2 \right) \\
\overline{\mathbf{w}} &= \begin{bmatrix} 1 \\ \overline{\mathbf{x}}_1 \\ \overline{\mathbf{x}}_2 \\ \overline{\mathbf{z}} \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{1}{n} \sum_{i=1}^n \mathbf{x}_{1i} \\ \frac{1}{m} \sum_{j=1}^m \mathbf{x}_{2j} \\ \frac{1}{N} \sum_{i=1}^N \mathbf{z}_i \end{bmatrix} \\
\widehat{\mathbf{V}}_{g_2} &= \frac{1}{m-1} \sum_{j=1}^m (\mathbf{x}_{2j} - \overline{\mathbf{x}}_{2j})(\mathbf{x}_{2j} - \overline{\mathbf{x}}_{2j})' - \widehat{\boldsymbol{\Sigma}}_2
\end{aligned}$$