



AMERICAN ASSOCIATION OF WINE ECONOMISTS

AAWE WORKING PAPER

No. 70

Economics

RETURN TO WINE:
A COMPARISON OF THE HEDONIC,
REPEAT SALES, AND HYBRID
APPROACHES

James J. Fogarty and Callum Jones

August 2010

www.wine-economics.org

Return to wine: A comparison of the hedonic, repeat sales, and hybrid approaches

James J Fogarty¹ and Callum Jones

Abstract. Comparisons between the return to wine and standard financial assets are complicated in that the return to wine must be estimated from infrequent sales of heterogeneous wine brands. Wine returns can be estimated using several different approaches, and here the performance of the hedonic approach, repeat sales approach, and hybrid approach are compared using 14,102 auction sale observations for Australian wine over the period 1988 to 2000. For the data set considered the results show that the hybrid approach provides the most efficient estimates, and that the repeat sales approach provides significantly higher total return estimates than the other two approaches. The portfolio diversification benefit attributed to holding wine is then shown to vary with estimation method.

Key Words: Return to wine, Price index

JEL: C33, G12

¹ Corresponding author: James Fogarty, School of Agricultural and Resource Economics, University of Western Australia, email: James.Fogarty@uwa.edu.au

1 Introduction

A variety of methods can be used to construct price indexes for infrequently traded heterogeneous goods such as art objects, premium wines, or houses. One popular approach is the hedonic price regression approach, and Triplett (2004) is a comprehensive and practical reference on hedonic price index construction. A key advantage of the hedonic price regression approach relative to other approaches is that all the available sales data can be used. There are however some limitations to the approach. For example, detailed information on good attributes is required and this may not always be available. Additionally, although no sales observations are discarded, the hedonic approach generally fails to make use of all the information contained in any given sales data set. Specifically, within any given sales data set there will be a subset of goods that sell more than once, and the hedonic price regression approach fails to take advantage of the additional information on repeat sales. In the specific context of the return to wine, the adjacent period hedonic price regression approach was used in Fogarty (2006) to estimate the return to Australian wine.

An alternative approach to the construction of a price index for infrequently traded heterogeneous goods is to consider only that subset of observations that sell more than once and construct a price index based on the repeat sales regression approach of Bailey et al. (1963). A key advantage of the repeat sales regression approach is that detailed information on good attributes is not required. The approach does, however, only use a subset of the total number of sales observations, and this is an issue if the subset of observations used is not representative of the full sample. Sample bias is potentially an issue for wine price indexes constructed using the repeat sales approach, for as noted in Ashenfelter et al. (1995), poor quality vintages tend to be traded less frequently than higher quality vintages. The repeat sales regression approach was used in Burton and Jacobsen (2001) and Sanning et al. (2008) to estimate the return to fine red Bordeaux wine; in Fogarty (2010) to estimate the return to Australian wine; in Masset and Weisskopf (2010) to estimate the return to Bordeaux,

Burgundy, Rhône Valley, Italian, and US wine; and a restricted form of the approach was used in Krasker (1979) and Jaeger (1981) to estimate the return to Red Bordeaux and California Cabernet.

The third approach to price index construction is the hybrid approach of Case and Quigley (1991). The hybrid approach seeks to combine features of both the hedonic and repeat sales approaches to achieve more efficient estimates of price change through time, and as yet this approach has not been used to estimate the return to wine. In the case of wine, where the proportion of observations that are repeat sales is generally high, and attributes are readily identifiable, it is not clear that point estimates and the associated standard errors from the hybrid approach will be noticeably different to those of the hedonic approach or the repeat sales approach.

To investigate whether estimation approach matters when constructing a wine price index the following paper compares estimates of the return to wine, and the associated standard errors, from the hedonic approach, the repeat sales approach, and the hybrid approach. As the existing return to wine literature is well summarised in both Sanning et al. (2008) and Fogarty (2010), an extensive literature review is not present here, and the remainder of the paper is structured as follows. Section 2 provides a unified notation that describes the various approaches that can be used to estimate the return to wine. Section 3 provides details on the data set used, presents comparative results, and uses a mean-variance spanning test to investigate whether the diversification benefit that has been attributed to wine is independent of estimation method. Section 4 provides concluding comments.

2 Methodology

Let $w \in \mathcal{W} = \{1, \dots, W\}$ be the set of observed wine sales, and let $t \in \mathcal{T} = \{0, 1, \dots, T\}$ be the set of time periods. Now, separate the set \mathcal{W} into the subset of wines that sell only once during

the sample period, $i \in \mathcal{I} = \{1, \dots, I\} \subset \mathcal{W}$ and the subset of wines that transact more than once, $j \in \mathcal{J} = \{1, \dots, J\} \subset \mathcal{W}$. With this notation the hedonic model can be written as:

$$p_{wt} = \alpha + \hat{\boldsymbol{\beta}}\mathbf{x}_w + \hat{\boldsymbol{\gamma}}\mathbf{d}_w + \eta_w + e_{wt}, \quad (1)$$

where p_{wt} is the log price of wine w sold at time t ; $\boldsymbol{\beta}$ is a $K \times 1$ vector of implicit prices for wine attributes; \mathbf{x}_w is a $K \times 1$ vector of attributes for wine w ; $\boldsymbol{\gamma}$ is a $T \times 1$ vector of index numbers; and \mathbf{d}_w is a $T \times 1$ vector of time dummies where the value at time t is one if wine w sold in time period t and zero otherwise; η_w is a zero mean specification error term assumed to be uncorrelated with \mathbf{x}_w and \mathbf{d}_w ; and e_{wt} is a random error term.

The repeat sales and hybrid models can be understood as follows. Re-write equation (1) so as to identify separately the repeat sale observations and the single sale observations, and for the subset of repeat sale observations let the first sale for wine j occur in time period t and the second sale occur in time period τ such that we have:

$$p_{it} = \alpha + \hat{\boldsymbol{\beta}}\mathbf{x}_i + \hat{\boldsymbol{\gamma}}\mathbf{d}_i + \eta_i + e_{it}, \quad (2a)$$

$$p_{jt} = \alpha + \hat{\boldsymbol{\beta}}\mathbf{x}_j + \hat{\boldsymbol{\gamma}}\mathbf{d}_j + \eta_j + e_{jt}, \quad (2b)$$

$$p_{j\tau} = \alpha + \hat{\boldsymbol{\beta}}\mathbf{x}_j + \hat{\boldsymbol{\gamma}}\mathbf{d}_j + \eta_j + e_{j\tau}. \quad (2c)$$

By differencing equation (2b) from equation (2c) the α , $\hat{\boldsymbol{\beta}}\mathbf{x}_j$, and η_j terms drop out to give the repeat sales regression price index:

$$p_{j\tau} - p_{jt} = \hat{\boldsymbol{\gamma}}\boldsymbol{\delta}_j + e_{j\tau} - e_{jt}, \quad (2d)$$

where $\boldsymbol{\delta}_j$ is a $T \times 1$ vector of time dummy variables taking the value minus one for the first sale, one for the second sale, and zero for all other cases.

The hybrid model is represented by equations (2a), (2b), and (2d), stacked in that order, and under the assumptions $E(\eta_w) = E(e_{wt}) = 0$, $\text{Var}(\eta_w) = \sigma_\eta^2$, $\text{Var}(e_{wt}) = \sigma_e^2$, and $\text{Cov}(e_{it}, e_{jt}) = \text{Cov}(e_{wt}, e_{w\tau}) = \text{Cov}(e_{wt}, \eta_w) = 0$ so that $\text{Var}(\varepsilon_{wt}) = \sigma_e^2 + \sigma_\eta^2$, the covariance matrix associated with the hybrid model can be written as:

$$\mathbf{\Omega} = \begin{bmatrix} \sigma_\varepsilon^2 \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_\varepsilon^2 \mathbf{I} & -\sigma_e^2 \mathbf{I} \\ \mathbf{0} & -\sigma_e^2 \mathbf{I} & 2\sigma_e^2 \mathbf{I} \end{bmatrix}. \quad (3)$$

Jones (2010) presents details on consistent estimators of σ_e^2 and σ_ε^2 . Specifically, $\hat{\sigma}_e^2$ is found by estimating regression equation (2d) on all J observations to obtain the residuals and then applying a degrees of freedom correction:

$$\hat{\sigma}_e^2 = \frac{1}{2} \left(\frac{1}{J-T} \right) \sum_{j=1}^J (\hat{\varepsilon}_{j\tau} - \hat{\varepsilon}_{jt})^2, \quad (4)$$

and $\hat{\sigma}_\varepsilon^2$ is found by estimating the regression shown at equation (1) across all W observations to obtain the residuals then and applying a degrees of freedom correction:

$$\hat{\sigma}_\varepsilon^2 = \left(\frac{1}{W-K-T-1} \right) \sum_{w=1}^W \hat{\varepsilon}_{wt}^2. \quad (5)$$

Once $\hat{\mathbf{\Omega}}$ has been found, the Cholesky decomposition can be used to find \mathbf{P} , where $\hat{\mathbf{P}}\mathbf{P} = \hat{\mathbf{\Omega}}^{-1}$, which in turn is used to transform the data prior to estimation via least squares. For a large data set finding $\hat{\mathbf{\Omega}}^{-1}$ directly is computationally intensive. However, note that the structure of $\hat{\mathbf{\Omega}}$ is such that there are only elements on the on-diagonal of each block, and that the non-zero values within each block are the same. As such, the inverse of a proportionally much smaller version of $\hat{\mathbf{\Omega}}$ can be found to obtain the values and location of the non-zero elements to create the full size $\hat{\mathbf{\Omega}}^{-1}$ and \mathbf{P} .

3 Data and results

The data were obtained from the Australian auction house Langton's, cover the period 1988 to 2000, and have been summarised at the quarterly frequency. Wine brands were only included in the sample if they were listed in Caillard and Langton (2001) as being of investment quality, and this resulted in consideration of 14,102 observations across 84 specific wine brands and 36 vintages. The 2001 edition of the Caillard and Langton wine investment guide has four quality tiers, and across the sample there were seven tier one wine brands, 20 tier two wine brands, 28 tier three wine brands, and 29 tier four wine brands, and summary information on the data set is provided in Table 1. From the first two columns of Table 1 it can be seen that the volume of Australian investment quality wine traded over the sample period grew substantially. Details on the main grape variety of the wines sold, and the quality rankings of the wines in the sample, can be read from the third and fourth columns of the table. As can be seen, most of the wines sold at auction were either Shiraz or Cabernet based wines, and although there were only seven tier one wine brands, these seven wine brands accounted for 22 percent of the sample. Details on observations classified by vintage are provided in the final six columns of Table 1, and it is notable that the most celebrated Australian vintage in recent decades was the 1990 vintage, and this vintage appeared most frequently in the sample.

Table 1 Description of the key features of the wine sales data set

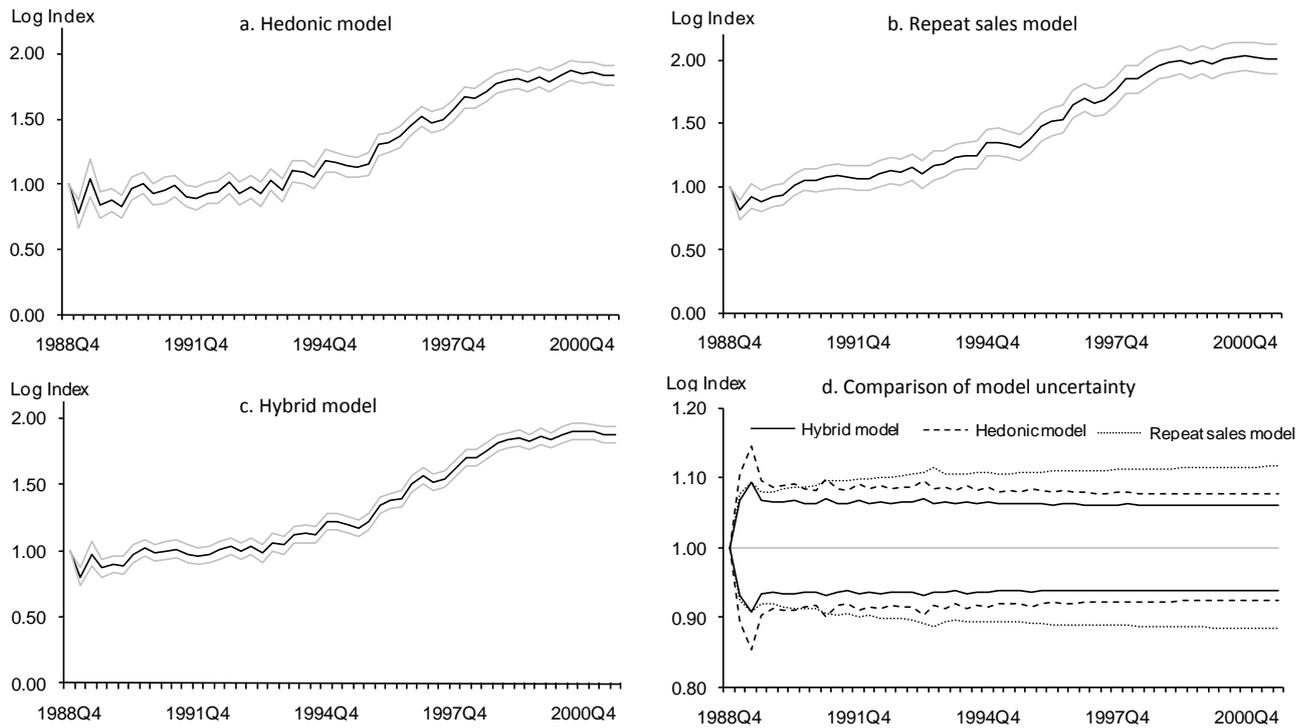
Sale year	Obs.	Wine type	Obs.	Vintage	Obs.	Vintage	Obs.	Vintage	Obs.
1988	160	Shiraz	5,356	1965	95	1978	313	1991	854
1989	418	Cabernet	6,086	1966	121	1979	337	1992	826
1990	560	Merlot	30	1967	119	1980	432	1993	719
1991	467	Pinot Noir	820	1968	105	1981	328	1994	685
1992	456	Botrytis	110	1969	118	1982	620	1995	483
1993	607	Chardonnay	1,192	1970	136	1983	412	1996	408
1994	785	Semillon	183	1971	169	1984	697	1997	193
1995	874	Riesling	325	1972	155	1985	739	1998	59
1996	1,253	<i>Quality Rating</i>		1973	184	1986	903	1999	14
1997	1,462	Tier 1	3,158	1974	157	1987	780	2000	1
1998	2,000	Tier 2	2,292	1975	173	1988	748		
1999	2,404	Tier 3	5,031	1976	254	1989	581		
2000	2,656	Tier 4	3,621	1977	236	1990	948		

The point estimates and standard errors for the price indexes estimated using each approach are shown in Table 2. As can be seen by considering the information in the respective standard error columns of Table 2, the hybrid model clearly provides the most efficient estimates. Specifically, across all time periods the average standard error for each model was as follows: hedonic model .074, repeat sales .053, and hybrid model .032. Figure 1 presents a visual summary of the information. In the first three panels of Figure 1 the solid black line represents the price index point estimate, and the grey bands represent the 95 percent confidence interval. To make the estimate efficiency comparison across models clearer, the final panel of Figure 1 plots the 95 percent confidence interval for each model around a constant index value of unity, and from the plot the relative efficiency of the hybrid approach can be seen clearly.

Table 2 **Log price index of wine prices**

	Hedonic model		Repeat sales model		Hybrid model	
	Estimate	SE	Estimate	SE	Estimate	SE
1988Q1	1.000	-	1.000	-	1.000	-
1988Q2	.778	(.054)	.817	(.039)	.803	(.035)
1988Q3	1.044	(.074)	.926	(.047)	.974	(.047)
1988Q4	.840	(.049)	.887	(.041)	.870	(.034)
1989Q1	.877	(.044)	.926	(.041)	.895	(.033)
1989Q2	.827	(.046)	.939	(.043)	.887	(.034)
1989Q3	.968	(.046)	1.016	(.044)	.976	(.034)
1989Q4	1.009	(.043)	1.056	(.044)	1.016	(.033)
1990Q1	.924	(.042)	1.049	(.045)	.981	(.032)
1990Q2	.957	(.050)	1.070	(.049)	.997	(.036)
1990Q3	.987	(.043)	1.086	(.049)	1.012	(.032)
1990Q4	.909	(.041)	1.076	(.048)	.976	(.032)
1991Q1	.896	(.046)	1.065	(.050)	.958	(.034)
1991Q2	.935	(.043)	1.065	(.050)	.974	(.033)
1991Q3	.937	(.045)	1.104	(.052)	1.002	(.033)
1991Q4	1.015	(.043)	1.128	(.052)	1.034	(.033)
1992Q1	.933	(.043)	1.114	(.052)	.997	(.033)
1992Q2	.983	(.044)	1.151	(.053)	1.033	(.033)
1992Q3	.926	(.049)	1.099	(.055)	.979	(.035)
1992Q4	1.034	(.043)	1.169	(.058)	1.061	(.033)
1993Q1	.957	(.044)	1.176	(.054)	1.039	(.033)
1993Q2	1.104	(.041)	1.226	(.054)	1.122	(.032)
1993Q3	1.093	(.045)	1.243	(.054)	1.128	(.033)
1993Q4	1.055	(.042)	1.248	(.054)	1.122	(.032)
1994Q1	1.180	(.044)	1.345	(.054)	1.221	(.033)
1994Q2	1.167	(.041)	1.354	(.054)	1.221	(.032)
1994Q3	1.139	(.041)	1.331	(.054)	1.189	(.032)
1994Q4	1.129	(.041)	1.310	(.054)	1.172	(.032)
1995Q1	1.156	(.043)	1.370	(.055)	1.219	(.033)
1995Q2	1.304	(.041)	1.473	(.055)	1.337	(.032)
1995Q3	1.321	(.040)	1.516	(.056)	1.372	(.031)
1995Q4	1.366	(.041)	1.532	(.056)	1.394	(.032)
1996Q1	1.444	(.041)	1.649	(.056)	1.500	(.032)
1996Q2	1.520	(.040)	1.700	(.056)	1.559	(.031)
1996Q3	1.475	(.040)	1.664	(.056)	1.518	(.031)
1996Q4	1.500	(.040)	1.682	(.057)	1.536	(.031)
1997Q1	1.566	(.040)	1.759	(.057)	1.610	(.031)
1997Q2	1.665	(.040)	1.848	(.057)	1.702	(.031)
1997Q3	1.659	(.039)	1.849	(.057)	1.700	(.031)
1997Q4	1.714	(.040)	1.903	(.057)	1.753	(.031)
1998Q1	1.773	(.039)	1.961	(.058)	1.812	(.031)
1998Q2	1.794	(.039)	1.977	(.058)	1.830	(.031)
1998Q3	1.811	(.039)	2.000	(.058)	1.852	(.031)
1998Q4	1.780	(.039)	1.965	(.058)	1.817	(.031)
1999Q1	1.819	(.039)	2.000	(.058)	1.856	(.031)
1999Q2	1.789	(.039)	1.973	(.058)	1.830	(.031)
1999Q3	1.834	(.039)	2.011	(.058)	1.873	(.031)
1999Q4	1.870	(.039)	2.023	(.059)	1.894	(.031)
2000Q1	1.853	(.039)	2.028	(.059)	1.895	(.031)
2000Q2	1.860	(.039)	2.025	(.059)	1.892	(.031)
2000Q3	1.834	(.039)	2.010	(.059)	1.875	(.031)
2000Q4	1.832	(.039)	2.002	(.059)	1.868	(.031)

Figure 1 Summary comparison of hybrid model, repeats sales model, pooled model and hybrid model



In addition to considering efficiency, there is the possibility that the repeat sales model suffers from sample selection bias and so may overstate the population return. From the log index values shown in the final row of Table 2 it can be seen that the hedonic model and the hybrid model give quite close total return estimates, but that the estimates from the repeat sales regression approach give a noticeably higher total return estimate. The use of a log price index does not necessarily make clear the extent of the difference between the return estimates from the repeat sales model and the other models, and as such it is helpful to consider the implied total return estimates for each model in percentage terms. For the hedonic model the estimated total return to wine for the period was 129.8 percent (standard error 4.0 percent); for the repeat sales model the estimated total return was 172.4 percent (standard error 6.1 percent); and for the hybrid model the estimated total return was 138.0 percent (standard error 3.2

percent).² As such, it appears that at least for Australian wine, use of the repeat sales approach results in an overstatement of the population return.

Both Fogarty (2010) and Masset and Weisskopf (2010) used the repeat sales regression methodology to estimate the return to wine and then show that a positive allocation to wine in a well diversified investment portfolio improves the risk-return profile of the investment portfolio. The results presented here suggest that the benefit attributed to wine may have been overstated due to the estimation method chosen in these studies. Although, if the benefit from wine is largely due to a lack of correlation between the return to wine and standard financial assets, even with a much lower mean return estimate, wine may still provide a portfolio diversification benefit. Table 3 provides information on the correlation between the return to wine and standard financial assets. The return to wine information shown is based on the return estimates from the hybrid model, and as can be seen by considering the correlation coefficient information in the first column of Table 3, the return to wine and standard financial assets are not strongly correlated. Correlation coefficient information alone is however not sufficient to determine whether or not wine provides a portfolio diversification benefit, and a formal test is required.

Table 3 **Pair-wise asset correlations 1988-2000**

Assets	Australian wine	Australian shares	Australian bonds	US shares	US Bonds
Australian wine	1.000				
Australian shares	.058	1.000			
Australian bonds	.031	.248	1.000		
US shares	.133	.400	.247	1.000	
US bonds	.170	-.101	.471	.568	1.000

A mean-variance spanning test is a formal test that can be used to investigate whether an asset provides a diversification benefit to an investment portfolio. When considering the addition of a single asset class to an existing investment portfolio an appropriate approach to

² In this case the impact of using the corrections proposed by Kennedy (1981) and Van Garderen and Shah (2002) for calculating the percentage return and the associated standard error are so small they can be safely ignored, and the formula $(\exp(X)-1) \times 100$ can be used.

test for mean-variance spanning is the Huberman and Kandel (1987) regression based test. In this instance the test involves regressing the return to wine on the return to the assets already in the investment portfolio and an intercept term. In such a regression if the intercept is zero, and the sum of the point estimates on the other asset classes is one, the conclusion drawn is that the return to wine can be mimicked by a weighted sum of the assets already in the investment portfolio, and the return to wine is said to be spanned by the existing assets. If the return to the test asset can be synthetically reproduced by the assets already in an investment portfolio the test asset is not added to the portfolio.

To test whether the finding that wine provides a diversification benefit is affected by the method used to calculate returns, the Huberman and Kandel mean-variance spanning regression was estimated using quarterly log return information for the period 1988 to 2000. The form of the spanning test regression was $r_{jt} = \alpha + \sum_{i=1}^K \beta_i r_{it} + u_t$, where r_{jt} denotes the return to wine from estimation method j at time t ; r_{it} denotes the return to benchmark asset i at time t , where the benchmark assets are the total returns to Australian shares and Australian bonds, and unhedged Australian dollar total returns to US shares, and US bonds; and u_t denotes a zero mean constant variance error term. Details for the mean-variance spanning regressions are shown in Table 4. For the repeat sales regression return estimates mean-variance spanning was strongly rejected, with the Wald test statistic p-value substantially less than one percent. Similarly, when using return estimates from the hybrid model the Wald test statistic p-value was substantially less than one percent so that the hypothesis of mean-variance spanning is also strongly rejected. Mean-variance spanning cannot, however, be rejected at conventional levels when using the hedonic model return to wine estimates. The diversification benefit attributed to wine is therefore not independent of the estimation method chosen.

Table 4 **Mean-variance spanning regression results for wine**

	Hedonic model		Repeat sales model		Hybrid model	
	Estimate	SE	Estimate	SE	Estimate	SE
Intercept	.008	(.017)	.012	(.009)	.010	(.012)
Australian shares	.277	(.279)	.019	(.151)	.125	(.191)
Australian shares	-.334	(.505)	-.112	(.273)	-.215	(.347)
US shares	-.084	(.234)	.022	(.126)	-.026	(.160)
US Bonds	.385	(.316)	.128	(.171)	.237	(.217)
Wald test p-value	.133		.0002		.005	

4 Conclusion

To date the main approach used to estimate the return to wine has been the repeat sales regression approach. Using a data set of Australian fine wine sales it has been shown that the hybrid approach due to Case and Quigley (1991) can be used to obtain more efficient estimates of the return to wine than the repeat sales approach, or the hedonic approach. This is because unlike the repeat sales approach the hybrid approach uses all of the available data, and unlike the hedonic approach it identifies repeat sales where they exist. Additionally, the study found some evidence that the repeat sales approach may overstate the return to wine. The diversification benefit attributed to wine was shown to be dependent on the estimation method employed. However, using the estimates from the preferred hybrid methodology, wine was shown to still provide a diversification benefit to an already well diversified investment portfolio.

5 References

- Ashenfelter, O., D. Ashmore and R. Lalonde. 1995. "Bordeaux Wine Vintage Quality and the Weather." *Chance*. 8: 7-14.
- Bailey, M.J., R.F. Muth and H.O. Nourse. 1963. "A Regression Method for Real Estate Price Index Construction." *Journal of the American Statistical Association*. 58: 933-942.
- Burton, J. and J.P. Jacobsen. 2001. "The Rate of Return on Wine Investment." *Economic Inquiry*. 39: 337-350.
- Case, B. and J. M. Quigley 1991. "The Dynamics of Real Estate Prices." *Review of Economics and Statistics*. Vol. 73, No. 1, pp. 50-58.
- Fogarty, J.J. 2006. "The Return to Australian Fine Wine." *European Review of Agricultural Economics*. 33: 542-561.
- Fogarty, J.J. 2010. "Wine Investment and Portfolio Diversification Gains." *Journal of Wine Economics*. 5: 1-13.
- Huberman, G. and S. Kandel. 1987. "Mean-Variance Spanning." *Journal of Finance*. 42: 873-888.
- Jaeger E. 1981. "To Save or Savor: The Rate of Return to Storing Wine: Comment." *Journal of Political Economy*. 89: 584-592.
- Jones C. 2010. "House Price Measurement: The Hybrid Hedonic Repeat-Sales Method." *Economic Record*. 86: 97-97.
- Kennedy, P. 1981. "Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations." *American Economic Review*. 71: 801.
- Krasker, W.S. 1979. "The Rate of Return to Storing Wine." *Journal of Political Economy*. 87: 1363-1367.
- Masset, P. and Henderson, C. (2010). Wine as an alternative asset class. *Journal of Wine Economics*, 5: forthcoming.
- Masset P. and Weisskopf, J. (2010). Raise your glass: wine investment and the financial crisis. American Association of Wine Economists Working Paper No. 57.
- Sanning, L.W., S. Shaffer and J.M. Sharratt. 2008. "Bordeaux Wine as a Financial Investment." *Journal of Wine Economics*. 3: 51-71.
- Triplett, J. 2004. Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products. STI Working Paper, OECD.
- Van Garderen, K. J. and S. Chandra. 2002. "Exact Interpretation of Dummy Variables in Semilogarithmic Equations." *Econometrics Journal*. 5:149-159.